

UNIVERZITA PALACKÉHO V OLOMOUCI
PŘÍRODOVĚDECKÁ FAKULTA

DIPLOMOVÁ PRÁCE

Statistická analýza časových řad klientských
vkladů v České republice



Katedra matematické analýzy a aplikací matematiky
Vedoucí diplomové práce: **doc. RNDr. Eva Fišerová, Ph.D.**
Vypracoval: **Bc. Jan Elgner**
Studijní program: N1103 Aplikovaná matematika
Studijní obor: Aplikace matematiky v ekonomii
Forma studia: prezenční
Rok odevzdání: 2017

BIBLIOGRAFICKÁ IDENTIFIKACE

Autor: Bc. Jan Elgner

Název práce: Statistická analýza časových řad klientských vkladů v České republice

Typ práce: Diplomová práce

Pracoviště: Katedra matematické analýzy a aplikací matematiky

Vedoucí práce: doc. RNDr. Eva Fišerová, Ph.D.

Rok obhajoby práce: 2018

Abstrakt: Diplomová práce se zabývá statistickou analýzou klientských vkladů registrovaných u České národní banky v období 1993 - 2016. V první a druhé části charakterizujeme vklady a teoretické aspekty statistické analýzy časových řad. Následně modelujeme datové soubory netermínovaných vkladů a krátkodobých termínovaných vkladů. Nakonec navržené modely porovnáme vzhledem k historickým a budoucím záznamům u obou datových souborů.

Klíčová slova: časové řady, statistická analýza, dekompoziční metody, klientské vklady, adaptivní metody

Počet stran: 74

Počet příloh: 4

Jazyk: český

BIBLIOGRAPHICAL IDENTIFICATION

Author: Bc. Jan Elgner

Title: Statistical analysis of client's deposits time series in the Czech Republic

Type of thesis: Master's

Department: Department of Mathematical Analysis and Application of Mathematics

Supervisor: doc. RNDr. Eva Fišerová, Ph.D.

The year of presentation: 2018

Abstract: The master's thesis deals with statistical analysis of client deposits registered with Czech National Bank in the period 1993 - 2016. In the first and second part, we characterize the deposits and the theoretical aspects of statistical analysis of time series. Afterwards, we create models for two data sets, namely non-term deposits and short term deposits. Finally, we compare proposed models with regard to historical and future records of both data sets.

Key words: time series, statistical analysis, decompose approach, deposits, adaptive methods

Number of pages: 74

Number of appendices: 4

Language: Czech

Prohlášení

Prohlašuji, že jsem diplomovou práci zpracoval samostatně pod vedením doc. RNDr. Evy Fišerové, Ph.D. a všechny použité zdroje jsem uvedl v seznamu literatury.

V Olomouci dne

.....

podpis

Obsah

Úvod	8
1 Vklady	9
1.1 Klasifikace vkladů	12
1.2 Volba investice v závislosti na časovém horizontu	13
2 Dekompoziční přístup k časovým řadám	14
2.1 Základní pojmy	14
2.2 Shrnutí přístupů k modelování ČŘ	15
2.3 Očištění dat	16
2.4 Dekompozice časové řady	17
2.4.1 Obecný lineární model	18
2.4.2 Trendová složka	24
2.4.3 Periodická složka	29
2.4.4 Adaptivní metody	34
2.4.5 Testy náhodnosti	44
3 Analýza klientských vkladů	49
3.1 Úprava a načtení dat	49
3.2 Porovnání trendových funkcí	54
3.3 Periodické modely	57
3.4 Adaptivní metody	61
3.5 Identifikace náhodné složky	65
Závěr	70
Literatura	72
A Trendová složka	75
A.1 Lineární trend	75
A.2 Kvadratický trend	78
A.3 Exponenciální trend	78
A.4 Logistický trend	80
B Periodická složka	81
B.1 Model s umělými proměnnými	81
B.2 Model skrytých period s kvadratickým trendem	82
C Adaptivní metody	86
C.1 Dvojitě exponenciální vyrovnávání	86
C.2 Holtova-Wintersova metoda	88

D Testy náhodnosti	90
D.1 Test založený na znaménkách diferencí	90
D.2 Test založený na bodech zvratu	90
D.3 Test založený na Kendallově koeficientu pořadové korelace τ . . .	91
D.4 Test založený na Spearmanově koeficientu ρ	92
D.5 Mediánový test	92

Poděkování

Na tomto místě bych rád poděkoval vedoucí mé diplomové práce doc. RNDr. Evě Fišerové, Ph.D. za čas věnovaný konzultacím, za inspirující nápady a za trpělivost. Velké poděkování patří také mojí rodině, přátelům a spolužákům, kteří mě neustále motivovali a podporovali po celou dobu studia.

Úvod

V běžné praxi se setkáváme s různými datovými soubory, jejichž hodnoty souvisí s časovým obdobím. Při setkání s daty tohoto typu tak vyvstávají otázky jako např. jaký vývoj datového souboru lze očekávat? Co je důvodem náhlých či dlouhodobějších změn od trendu? Jaké modely jsou vhodné pro zjištění těchto okolností? Na všechny otázky se budeme snažit odpovědět v následujícím textu, ve kterém budeme analyzovat vývoj klientských vkladů. Datový soubor těchto vkladů pochází z databáze České národní banky a jedná se o veškeré uložené vklady v komerčních bankách spravované českou centrální bankou v období 1993 - 2016.

Práce je složena ze tří částí. Pro lepší orientaci v problematice, vztahující se k datovému souboru, uvedeme nejprve ekonomické souvislosti klientských vkladů.

V druhé části se seznámíme s teoretickými aspekty dekompozičního přístupu časových řad. Kromě modelování trendové složky se budeme zabývat modelováním sezónní, cyklické a náhodné složky, a to pomocí trendových funkcí, modelu s umělými proměnnými, modelu skrytých period, adaptivních metod a testů náhodnosti.

Následuje stěžejní část této práce, ve které analyzujeme časové řady klientských vkladů v České republice. Datové soubory očistíme od výše cenové hladiny a kalendářních dní. Dále aplikujeme metody popsané v druhé části.

Na samotný závěr představíme charakteristiky jednotlivých modelů a vybereme z nich ty nejvhodnější, které následně doplníme grafy s bodovými a intervalovými odhady. Zároveň zmíníme možné ekonomické souvislosti s již dosaženými výsledky z předchozí části práce. Veškeré výpočty jsou provedeny ve statistickém softwaru R. Odpovídající kódy můžeme nalézt v přílohách této práce a na příloženém CD.

1. Vklady

V této kapitole se zaměříme na problematiku vkladů a jejich využití v praxi. Vklady můžeme chápat jako nástroj pro uspokojování budoucích potřeb. Jsou nedílnou součástí každé ekonomiky a prostřednictvím bank a jiných finančních organizací podněcují investiční činnost. Často se setkáváme s termínem depozitum, jež vyjadřuje výši vkladu v peněžních jednotkách. V této kapitole vycházím z [1] a [2].

Definice 1.1. *Vklady* jsou svěřené peněžní prostředky, které představují závazek finanční instituce vůči vkladateli. Zahrnují konečné zůstatky vkladových účtů klientů včetně připsaných úroků, tj. součet počátečních zůstatků a uložených depozit, od kterých odečteme vybraná depozita. Ve stavech vkladů jsou zahrnuty i vklady vládních institucí.

Na trhu se můžeme setkat se dvěma typy ekonomických subjektů. Na jedné straně jsou domácnosti, jež poskytují své finanční zdroje finančním organizacím za účelem jejich zhodnocení. Vklady klientů přitom vznikají vytvářením úspor volných peněžních prostředků, které nebyly použity ke spotřebě. Na druhé straně se setkáváme s podnikatelskými subjekty, k nimž jsou finanční zdroje přesunovány formou poskytovaných úvěrů. Tyto vazby představují jednu ze základních možností, jak využívat finanční zdroje (kapitál), které by jinak ztrácely svoji hodnotu, např. vlivem inflace.

Nyní si ujasníme, jaké možnosti vkladů mají klienti, kteří chtějí své peníze uložit a nejlépe zhodnotit. Těmito klienty mohou být fyzické nebo právnické osoby, ať už z tuzemska, či ze zahraničí, a dále stát. Je možné využít některé z níže uvedených zprostředkovatelů a jejich produktů:

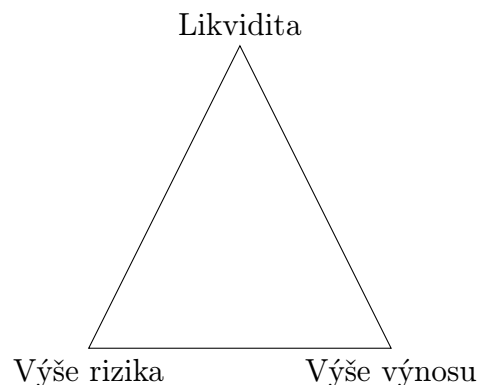
1. banky a spořitelny - např. bankovní a spořicí účty, stavební spoření,
2. pojišťovny - např. kapitálové a investiční životní pojištění,
3. investiční společnosti a fondy,

4. burza, RM-S¹, finanční brokerské firmy.

Vyvstává otázka, jak zprostředkovatelé nabytý kapitál od klientů zhodnotí. Pro finanční instituce jsou vklady od klientů velmi výhodné, neboť jim je za jejich vklad nabízen mnohem menší úrok než v opačném případě. Ten pro finanční instituci představuje část nákladů. Naopak výnosy získávají tyto instituce z úroků za poskytnuté úvěry. Tyto úroky jsou podstatně vyšší a instituce z nich pak hradí své náklady, a to jak klientům, tak na provoz své organizace či na rezervy.

Vklady se tak stávají oboustranně prospěšné, jak pro vlastníka, tak pro zprostředkovatele. Motivací obou stran je zhodnocení finančních prostředků a dosažení zisku. Nicméně, musíme si uvědomit, že každá investice obnáší riziko, jehož výši lze ilustrovat pomocí investičního trojúhelníku (obrázek 1). Můžou tak nastat tři situace, které odrážejí investici:

- vysoký výnos a vysoká likvidita
- vysoký výnos a nízká výše rizika
- vysoká likvidita a nízká výše rizika



Obrázek 1: Investiční trojúhelník.

¹Jedná se o burzu cenných papírů, kde se obchoduje s akciemi českých a zahraničních firem.

Provázanost všech tří vrcholů nelze obejít. Proto je potřeba hledat takové produkty a možnosti investic, které uspokojují naše potřeby a také naši toleranci k riziku.

Zhodnocení vkladů je umožněno díky finančnímu trhu, na kterém se střetává nabídka a poptávka po finančních prostředcích. Bez trhu by nemohlo k žádné směně ani investici dojít. Přitom poptávka po volných finančních prostředcích plyne z potřeby mít k dispozici prostředek směny - peníze.

Základem zhodnocení peněz je jejich obrat (přeměna peněz v kapitál a jejich následný výnos), neboť držením peněz přicházíme o možné výnosy. Rozhodnutí o držbě či investici závisí na preferencích klienta.

Zhodnocení, neboli úrok vkladu, je možné chápat jako jeho cenu. Výše úroků závisí zejména na výši poptávky a nabídky po finančních zdrojích. Díky velké konkurenci na trzích se cena vkladů utváří velmi reálně a je nabízena široká škála produktů s rozdílnými úroky. Nyní si popíšeme druhy bankovních depozitních obchodů.

Přímé depozitní obchody - jedná se o prvotní vklady (neboli primární depozita) od fyzických a právnických osob, které ještě nebyly uloženy v žádné jiné bance. Pro banku je jejich získávání velmi obtížné a pro jejich objevení je potřeba vytvořit vzájemnou důvěru mezi klientem a finanční institucí (bankéřem, finančním poradcem).

Nepřímé depozitní obchody - týká se bank, či peněžních ústavů, které poskytují své vklady finanční instituci.

Depozitní obchody s ČNB - jde o oboustranné přesuny depozit v rámci finanční instituce a ČNB.

Četnost jednotlivých obchodů pak ovlivňuje cena vkladů a jejich likvidita. Každá finanční instituce tak zvažuje, zda za vklady bude platit primárnímu klientovi, u kterého ale nastává riziko, že finanční zdroje vyčerpá, nebo je získá na mezibankovním trhu, případně od ČNB. Každá finanční instituce se následně rozhoduje dle svých preferencí a potřeb. Zejména posuzuje, co je pro ni z finančního hlediska nejvýhodnější. Obecně platí, že primární depozita jsou pro finanční

instituci levnější než nepřímé depozitní obchody a depozitní obchody s ČNB.

1.1. Klasifikace vkladů

Vklady je možné dělit dle různých hledisek, např. dle ekonomických sektorů, časového hlediska, či druhového hlediska.

Dle ekonomických sektorů rozlišujeme vklady podle institucí, ze kterých pochází. Jedná se o nefinanční instituce, finanční podniky, vládní instituce, domácnosti, neziskové instituce sloužící domácnostem a nerezidenty².

Z časového hlediska dělíme vklady na netermínované (bez uvedení doby splatnosti/výpovědi), na krátkodobé s dobou splatnosti (výpovědi) do jednoho roku včetně, na střednědobé s dobou splatnosti (výpovědi) do pěti let včetně a dlouhodobé s dobou splatnosti (výpovědi) nad pět let.

Dle druhového hlediska rozlišujeme vklady netermínované, termínované a podle skupin bank. Netermínované vklady se týkají běžných, kontokorentních účtů, netermínovaných účelových a jiných vkladů, úsporných vkladů na požádání (např. vklad na vkladní knížce) a také na netermínované vklady ve stavebním spoření. Termínované vklady zahrnují vkladové certifikáty, termínované vklady se splatností s výpovědní lhůtou, úsporné vklady se splatností s výpovědní lhůtou, vklady ve stavebním spoření se splatností s výpovědní lhůtou, účelově vázané vklady se splatností s výpovědní lhůtou, podřízené vklady, podřízené úvěry (přijaté), úvěry z repo obchodů (přijaté) a státem přijaté úvěry. Podle skupin banky rozdělujeme na základě velikosti bilanční sumy s doplněním o organizační a specializační aspekty. Rozdělení bank je kontrolováno vždy na začátku kalendářního roku a v případě potřeby se upravují hraniční body intervalů pro rozdělení bank. Banky klasifikujeme na malé, střední a velké (tabulka 1). Mezi velké banky tak řadíme např. Českou spořitelnu, Československou obchodní banku nebo Komerční banku. Zdroj výše aktiv těchto bank byl zjištěn z [18, 19, 20].

²Jedná se o daňového rezidenta, jenž má omezenou daňovou povinnost a platí pouze za příjmy z tuzemských zdrojů [3].

Tabulka 1: Klasifikace bank dle výše aktiv.

Období	Aktiva v mld. Kč		
	Malá	Střední	Velká
2002 až 2006	do 20	20 až 100	nad 100
2007 až 2008	do 50	50 až 150	nad 150
2009 až 2012	do 50	50 až 200	nad 200
2013 až 2015	do 50	50 až 250	nad 250
2016 a později	do 100	100 až 400	nad 400

1.2. Volba investice v závislosti na časovém horizontu

Obecně pro krátkodobější investice platí, že bychom měli využívat méně rizikových produktů. Na delší časový horizont si můžeme dovolit využít rizikovější produkty, přičemž důležitá je diverzifikace našich vkladů do jednotlivých produktů. Čím je rozmanitost jednotlivých produktů větší, tím více snižujeme riziko naší celkové investice. Za nejbezpečnější považujeme případy, kdy jsou jednotlivé produkty navzájem negativně korelované, což znamená, že při poklesu jedné investice má druhá investice tendenci růst a naopak. Investici považujeme za úspěšnou, pokud zhodnocení našich kladů je vyšší než úroveň inflace za totéž období.

Nyní si popíšeme typy produktů, jenž jsou vhodné pro vklady v závislosti na délce investice. Pro krátkodobé investice se hodí zejména netermínované produkty, jelikož odráží vysokou likviditu. V praxi to znamená, že vklady máme kdykoli k dispozici. Pro dlouhodobější investice můžeme využít termínované produkty nebo ještě rizikovější podílové fondy či investice do cenných papírů. Delší dobu investice odráží i možné vyšší úroky.

Při výběru typu produktů bychom se měli rozhodovat na základě výše čistého úroku (úroku po zdanění). Mějme na paměti, že úroky z veškerých bankovních, pojistných a státních produktů (např. investiční životní pojištění, státní dluhopisy) je nutné zdanit srážkovou daní ve výši 15%. Naopak příjmy z prodeje cenných papírů a podílových fondů jsou osvobozeny od daně dle paragrafu 4 zákona o daních z příjmů [4]. Z dlouhodobějšího hlediska je vhodnější nechat sestavení diverzifikovaného portfolia z cenných papírů či podílových fondů na finančním poradci.

2. Dekompoziční přístup k časovým řadám

Ve statistice máme k dispozici celou řadu metod pro modelování časových řad (dále ČŘ). Mezi běžně vyhledávané patří přístup Boxe-Jenkinse, který se zaměřuje na modelování náhodné složky. Jelikož se však budeme zabývat ČŘ, která má již dle vizualizace lineární charakter, modelování pomocí modelů Boxe-Jenkinse by bylo zbytečně složité.

Naším cílem je nalézt co nejjednodušší model, u kterého budeme klást velký důraz na kvalitu předpovědi. Pro tento cíl nám postačí rozložení ČŘ pomocí dekompozičního přístupu. V rámci této kapitoly vycházím z [5, 6, 7, 8].

2.1. Základní pojmy

Před samotnou dekompozicí ČŘ je nutné uvést základní pojmy, bez kterých se v dalším textu neobejdeme.

Definice 2.1. *Náhodný proces* je systém náhodných veličin, které jsou definované na stejném pravděpodobnostním prostoru.

Náhodný proces označujeme jako Y_t , $t \in T$, kde T je indexová množina, pro případ ČŘ se jedná o čas. Dle indexové množiny T lze náhodný proces rozlišit na náhodnou posloupnost, náhodný proces a náhodnou funkci.

O náhodnou posloupnost se jedná, pokud T je spočetná množina, např. $T = \mathbb{Z}$, nebo $T = \mathbb{N}$, což je speciální případ náhodného procesu s diskretním časem. Náhodný proces uvažujeme v případě, že T je interval, např. $T \in \langle a, b \rangle$, nebo $T \in \langle 0, \infty \rangle$, což je speciální případ NP se spojitým časem. Poslední možností je situace, kdy T je vícerozměrná množina, např. $T = \mathbb{R}^k$, $k \geq 2$. Pak se jedná o náhodnou funkci.

Definice 2.2. *Časovou řadou (ČŘ)* rozumíme

- a) náhodnou posloupnost s diskretním časem;
- b) diskretizaci náhodného procesu se spojitým časem, tj. náhodný proces v okamžicích, kde $t_1 < t_2 < t_3 < \dots$;

c) konečnou verzi a) nebo b).

ČŘ můžeme vnímat jako posloupnost srovnatelných pozorování.

Nyní si uvedeme základní dělení ČŘ. ČŘ rozlišujeme na ekvidistantní a neekvidistantní. U ekvidistantních ČŘ jsou intervaly mezi pozorováními stejně dlouhé, tj. $(t = 1, 2, 3, \dots)$. Naproti tomu u neekvidistantních ČŘ se jednotlivé časové intervaly svou délkou liší, tj. $(t = t_1, t_2, t_3, \dots)$.

ČŘ dále dělíme na intervalové a okamžikové. Intervalová ČŘ vzniká nahromaděním dat za dané časové období, tedy výše hodnot závisí na délce intervalu. V intervalových datech je možné tvořit součty. Jedná se např. o počty nehod, výrobků, příjmy na bankovním účtu, apod. Naopak okamžiková ČŘ vzniká diskretizací spojitě veličiny, nebo se jedná o diskrétní ČŘ z podstaty věci. Hodnoty se často vztahují ke konkrétnímu období. V okamžikových ČŘ součty nedávají smysl, ale lze např. počítat průměry. Může se jednat o měření teploty, výše ceny statků a služeb, konečné zůstatky peněz na bankovním účtu, apod.

Posledním dělením ČŘ, které uvedeme, je dělení dle frekvence pozorování na krátkodobé nebo dlouhodobé ČŘ. U krátkodobých ČŘ jsou okamžiky mezi pozorováními kratší než jeden rok (jedná se o hodiny, dny, týdny, měsíce, ...), u dlouhodobých ČŘ je rozmezí mezi hodnotami ČŘ alespoň roční.

2.2. Shrnutí přístupů k modelování ČŘ

Obecný model ČŘ uvažujeme ve tvaru

$$y_t = f(t) + e_t,$$

kde y_t odpovídají jednotlivým pozorováním ČŘ v čase t , f je nenáhodná funkce a e_t označuje náhodnou složku.

Dekompoziční model se zaměřuje na modelování nenáhodné funkce $f(t)$, související s nenáhodnými složkami ČŘ, přičemž předpokládáme nekorelovanost náhodných chyb. ČŘ můžeme rozložit na čtyři složky, které se snažíme rozlišit a modelovat. Jedná se o trendovou, sezónní, cyklickou a náhodnou složku.

Velmi užitečným nástrojem pro modelování periodické složky (sezónní a cyklické složky) je **spektrální analýza**. V tomto přístupu je ČŘ považována za směs sinusových a kosinusových křivek s různými amplitudami a frekvencemi. **Boxova-Jenkinsova metodologie** spočívá v modelování náhodné složky ČŘ, kde předpokládáme, že chyby jsou korelované. ČŘ tak chápeme jako posloupnost závislých náhodných veličin a pomocí korelační analýzy se zkoumají tyto závislosti. Ve statistice je velmi cenné, pokud se nám povede zachytit příčinnou závislost. Touto problematikou se zabývají **příčinné modely**. Nicméně, příčinná závislost bývá doprovázena velmi náročnými modely, které nazýváme též kauzální nebo faktorové modely. Mezi příčinné modely řadíme i vícerozměrné modely.

2.3. Očištění dat

Před samotnou dekompozicí je velmi důležité provést očištění ČŘ od nepravidelných vlivů, které by mohly zkreslit výsledky naší analýzy. Mezi tyto vlivy můžeme řadit jakékoliv změny dat v čase, jako např. nepravidelné délky měsíců, vliv pracovních dnů a svátků, nepravidelná výše zhodnocení, či různé výše zaplacených daní za sledovaný časový úsek.

Nyní si popíšeme vhodné postupy, jak ČŘ očistit. Začneme očištěním od kalendářních dnů. Očištění od kalendářních dnů si nejlépe představíme na případě, kdy první den v měsíci je státní svátek. Lidé mají tendenci se předzásobit ještě v předchozím měsíci, což má za následek zvýšení tržeb v předchozím měsíci. Ony nepravidelnosti můžeme potlačit aplikací následujícího vztahu:

$$y_t^{(o)} = y_t \cdot \frac{\bar{k}_t}{k_t}, \quad (2.1)$$

kde y_t značí původní hodnotu očišťované časové řady, k_t je počet kalendářních dnů v daném období a \bar{k}_t odpovídá průměrnému počtu dní v tomtéž období. Pokud by dané období mělo délku jednoho měsíce, pak hodnotu \bar{k}_t zvolíme ve výši 30, nebo přesněji $365/12 = 30,41\bar{6}$, případně $366/12 = 30,5$, pokud se jedná o přestupný rok.

Dále se můžeme setkat s očištěním od pracovních dnů, které je velmi podobné předchozímu případu. Pro úpravu dat použijeme vztah

$$y_t^{(o)} = y_t \cdot \frac{\bar{p}_t}{p_t}, \quad (2.2)$$

kde symbolem p_t značíme počet pracovních dnů v daném období a \bar{p}_t značí průměrný počet pracovních dní v daném období.

Posledním typem očištění, o kterém se zmíníme, je očištění od nepravidelného zhodnocování. V tomto případě nás zajímají různé výše úroků nebo inflace za sledované časové období. V případě, že budeme chtít hodnoty ČŘ převést ke konci sledovaného období, využijeme úročení prostřednictvím vztahu

$$y'_t = y_t \prod_{i=t}^n (1 + \alpha_i). \quad (2.3)$$

Pokud potřebujeme převést hodnoty ČŘ na začátek sledovaného období, využijeme diskontování:

$$y'_t = y_t \prod_{i=1}^t (1 - \alpha_i)^{-1}, \quad (2.4)$$

kde t indexuje v obou případech aktuální období, n je rozsah výběru, tedy celkový počet sledovaných období časové řady a α_i představuje míru zhodnocení k i -tému období, např. míru inflace.

2.4. Dekompozice časové řady

Jak bylo uvedeno v podkapitole 2.1, ČŘ můžeme rozložit na čtyři složky. Jedná se o trendovou složku (Tr_t), popisující dlouhodobé změny ČŘ, sezónní složku (S_t), vztahující se ke krátkodobým výkyvům od trendu, cyklickou složku (C_t), zabývající se dlouhodobými výkyvy od trendu, a reziduální (zbytkovou) složku (E_t). Sezónní a cyklická složka dohromady tvoří složku periodickou. Jak si dále popíšeme, je velmi zajímavé zkoumat každou složku zvlášť, tedy izolovat je. Výhoda přístupu tkví v možnosti srovnání jednotlivých složek stejného typu

u různých ČŘ. Naším cílem je nejen zkoumání historických dat dané ČŘ, ale zajímá nás i její extrapolace, neboli vývoj ČŘ do budoucnosti. Neméně zajímavá je i oblast očištění ČŘ od jednotlivých složek. Může nás zajímat např. pouze trendová a náhodná složka, což vede k tomu, že ČŘ očistíme od periodické složky. Co se týče dekompozice, rozlišujeme dva typy rozkladů.

V případě **aditivního rozkladu** je ČŘ ve tvaru $y_t = Tr_t + S_t + C_t + E_t$. Tehdy uvažujeme každou složku ve skutečných hodnotách a jednotku každé složky volíme v souladu s y_t .

Dalším typem je **multiplikativní rozklad**, kdy ČŘ zapisujeme ve tvaru $y_t = Tr_t \cdot S_t \cdot C_t \cdot E_t$. V tomto případě uvažujeme ve skutečných hodnotách jen trendovou složku (Tr_t). Ostatní složky popisujeme v relativních hodnotách vzhledem k trendové složce. Multiplikativní rozklad lze i vhodně převést na aditivní rozklad pomocí logaritmické transformace.

Výhodou těchto rozkladů je, že nemusíme brát v úvahu všechny složky, ale jen některé. Pokud bychom uvažovali např. $y_t = Tr_t + E_t$, tak v tomto modelu vynecháváme periodickou složku. Než se seznámíme s analýzou jednotlivých složek, uvedeme obecný lineární model.

2.4.1. Obecný lineární model

Obecný lineární model uvažujeme ve tvaru

$$\underbrace{y}_{n \times 1} = \underbrace{X}_{n \times p} \underbrace{\beta}_{p \times 1} + \underbrace{\epsilon}_{n \times 1}, \quad (2.5)$$

kde y označuje vektor pozorování, matice X je matice plánu, jejíž sloupce se vztahují k jednotlivým vysvětlujícím proměnným, symbolem β rozumíme vektor neznámých parametrů a ϵ je vektor chyb. Jednotlivé vektory chápeme jako sloupcové.

Odhady parametrů v lineárním modelu získáme metodou nejmenších čtverců (dále MNČ). Musíme si však uvědomit, že ačkoliv MNČ poskytuje nejlepší nestranné lineární odhady, tak má své předpoklady a omezení (nekorelovanost chyb,

náchylnost na odlehlá pozorování nebo problémy při lineární závislosti vysvětlujících proměnných). V případě, že bychom chtěli odhadnout parametry nelineárních modelů, postupujeme tak, že je převedeme na lineární modely, a to pomocí vhodně zvolené transformace.

Nyní uvedeme předpoklady, které jsou nezbytné, abychom model považovali za regulární. Jedná se o tyto předpoklady:

1. $n > p$, tj. rozsah výběru je větší než počet parametrů,
2. $h(X) = p$, tj. matice X má plnou sloupcovou hodnost,
3. $E(\epsilon_i) = 0$, $\text{var}(\epsilon_i) = \sigma^2$, pro $i = 1, \dots, n$,
 $\text{cov}(\epsilon_i, \epsilon_j) = 0$, pro $i \neq j$, tj. nekorelovanost chyb.

Třetí předpoklad požaduje, aby chyby ϵ tvořily bílý šum $\epsilon \sim WN(0, \sigma^2 I_n)$, kde WN značí white noise.

MNČ odhad parametrů modelu (2.5) lze vypočítat prostřednictvím vztahu

$$\hat{\beta} = (X^T X)^{-1} X^T y, \quad (2.6)$$

přičemž přesnost odhadu $\hat{\beta}$ můžeme vyjádřit varianční maticí:

$$\text{var}(\hat{\beta}) = \text{var}((X^T X)^{-1} X^T y) = (X^T X)^{-1} X^T \sigma^2 I_n X (X^T X)^{-1} = \sigma^2 (X^T X)^{-1}. \quad (2.7)$$

V dalším textu uvažujme počet pozorování n , počet předpovědí m , dále počet regresorů k a celkový počet parametrů p . Odhady budeme značit symbolem „ $\hat{\cdot}$ “. V případě, že budeme uvažovat model s absolutním členem, tj. první sloupec matice plánu X bude odpovídat jednotkovému vektoru, tak $p = k + 1$. Pokud model uvažujeme bez absolutního členu, tak $p = k$.

Regularita modelu je důležitá např. kvůli možnosti využití MNČ, případně intervalů spolehlivosti ve standardním tvaru. Porušení předpokladů regularity z důvodu heteroskedasticity chyb vede na zobecněnou MNČ, a tudíž i na zobecněné intervaly spolehlivosti. V případě, že by chyby byly navíc normálně rozdělené, jednalo by se o regulární lineární model, tj. $y \sim \mathcal{N}(X\beta, \sigma^2 I_n)$.

V rámci modelu se zjišťují vyrovnané hodnoty pro období $t = 1, \dots, n$, které získáme vztahem

$$\hat{y} = X\hat{\beta}, \quad (2.8)$$

Meze $(1 - \alpha) \cdot 100\%$ oboustranného intervalu spolehlivosti jsou pro očekávané hodnoty v čase t , resp. pro celý vektor rovny

$$\hat{y}_t \pm t_{(n-p), 1-\frac{\alpha}{2}} \hat{\sigma} \sqrt{(x_t^T (X^T X)^{-1} x_t)}, t = 1, \dots, n, \quad (2.9)$$

resp.

$$\hat{y}_t \pm \hat{\sigma} \sqrt{p \cdot F_{(p, n-p), 1-\frac{\alpha}{2}} (x_t^T (X^T X)^{-1} x_t)}, t = 1, \dots, n. \quad (2.10)$$

kde α představuje hladinu významnosti, označením $t_{(n-p), 1-\frac{\alpha}{2}}$, resp. $F_{(p, n-p), 1-\frac{\alpha}{2}}$ rozumíme $(1 - \frac{\alpha}{2})$ kvantil Studentova, resp. Fisherova rozdělení o příslušném počtu stupňů volnosti, x_t^T je t -tý řádkový vektor o rozměru $1 \times p$ matice X . Předpis (2.10) označujeme termínem simultánní interval spolehlivosti neboli pás spolehlivosti.

Bodové předpovědi pro čas $t = n + 1, \dots, n + m$ získáme přenásobením předpovědní datové matice $X_{(p)}$ s odhadem vektoru parametrů dle vztahu

$$\hat{y}^{(p)} = X_{(p)}\hat{\beta}, \quad (2.11)$$

kde předpovědní datová matice $X_{(p)}$ je rovna datové matici X pro pozorování v časech $n + 1, \dots, n + m$. Celkový počet budoucích hodnot je roven m . Označení (p) pak zdůrazňuje spojitost s budoucími pozorováními.

Obdobně jako v předchozím případě, můžeme zkonstruovat meze $(1 - \alpha) \cdot 100\%$ předpovědního oboustranného intervalu pro předpovídané hodnoty předpisem

$$\hat{y}_t^{(p)} \pm t_{(n-p), 1-\frac{\alpha}{2}} \hat{\sigma} \sqrt{1 + (x_t^T (X^T X)^{-1} x_t)}, t = n + 1, \dots, n + m, \quad (2.12)$$

nebo pro celý vektor (simultánní předpovědní interval) aplikací

$$\hat{y}_t^{(p)} \pm \hat{\sigma} \sqrt{m \cdot F_{(m, n-p), 1-\frac{\alpha}{2}} (1 + x_t^T (X^T X)^{-1} x_t)}, t = n + 1, \dots, n + m. \quad (2.13)$$

Uvedené intervalové odhady uvažujeme za předpokladu normality datového souboru. Kromě vyrovnaných hodnot, předpovědí a oboustranných intervalů nás v rámci modelu zajímají vybrané charakteristiky ČŘ. Jedná se zejména o: rezidua

$$\hat{e}_t = y_t - \hat{y}_t, t = 1, \dots, n, \quad (2.14)$$

reziduální součet čtverců

$$\text{RSČ} = \hat{e}^T \hat{e} = \sum_{t=1}^n (\hat{e}_t)^2 = \sum_{t=1}^n (y_t - \hat{y}_t)^2, \quad (2.15)$$

reziduální rozptyl

$$\hat{\sigma}^2 = \frac{\text{RSČ}}{n - p} \quad (2.16)$$

a směrodatnou odchylku

$$\hat{\sigma} = \sqrt{\hat{\sigma}^2}. \quad (2.17)$$

K posouzení vhodnosti modelu se využívá index determinace ve tvaru

$$R^2 = 1 - \frac{\text{RSČ}}{S_t^2}, \quad (2.18)$$

přičemž symbol S_t^2 představuje celkový součet čtverců. V modelu s absolutním členem jej vyjádříme vztahem

$$S_t^2 = (y - \bar{y})^T (y - \bar{y}) = \sum_{t=1}^n (y_t - \bar{y})^2, \quad (2.19)$$

v modelu bez absolutního členu se výpočet celkového součtu čtverců zjednoduší na

$$S_t^2 = y^T y = \sum_{t=1}^n y_t^2. \quad (2.20)$$

Jelikož index determinace nezohledňuje počet parametrů v modelu, tak pro menší rozsah datového souboru je vhodnější použít upravený index determinace,

který zohledňuje počet parametrů v modelu. Upravený index determinace vyjádříme vztahem

$$\begin{aligned} R_a^2 &= 1 - \frac{(n-1) \text{RSČ}}{(n-p) S_t^2} \\ &= 1 - (1 - R^2) \frac{n-1}{n-p}. \end{aligned} \quad (2.21)$$

Oba indexy determinace (2.18) a (2.21) nabývají hodnot z intervalu $\langle 0, 1 \rangle$. Index determinace vyjadřuje míru variability, jež je modelem vysvětlena. Rostoucí index determinace popisuje kvalitnější vztah modelu k historickým pozorováním datového souboru.

Mezi další charakteristiky, jež popisují kvalitu modelu z hlediska historických dat, můžeme zařadit střední čtvercovou chybu, kterou vyjádříme vztahem

$$\text{MSE} = \frac{\text{RSČ}}{n}. \quad (2.22)$$

Abychom mohli střední čtvercovou chybu porovnat se střední absolutní odchylkou, musíme si ze střední čtvercové chyby vyjádřit odmocninu

$$\text{RMSE} = \sqrt{\frac{\text{RSČ}}{n}}. \quad (2.23)$$

Střední absolutní odchylka je ve tvaru

$$\text{MAD} = \frac{1}{n} \sum_{t=1}^n \left| \hat{e}_t^{(p)} \right|. \quad (2.24)$$

Ve prospěch modelu, vzhledem k historickým hodnotám datového souboru, svědčí nižší hodnoty charakteristik (2.22), (2.23) a (2.24).

Kvalitu modelu z hlediska předpovědi popisujeme prostřednictvím následujících charakteristik, které vychází z budoucích hodnot datového souboru. Mezi tyto charakteristiky řadíme součet čtvercové chyby předpovědi vyjádřený vztahem

$$\text{SEP} = \sum_{t=n+1}^{n+m} (\hat{e}_t^{(p)} - \bar{e}^{(p)})^2, \quad (2.25)$$

kde $\bar{e}^{(p)}$ je aritmetický průměr reziduí z období $t = n+1, \dots, n+m$. Dále střední čtvercovou chybu předpovědi ve tvaru

$$\text{MSEP} = \frac{1}{m} \sum_{t=n+1}^{n+m} (y_t - \hat{y}_t^{(p)})^2, \quad (2.26)$$

odmocninu ze střední čtvercové chyby

$$\text{RMSEP} = \sqrt{\frac{1}{m} \sum_{t=n+1}^{n+m} (y_t - \hat{y}_t^{(p)})^2}, \quad (2.27)$$

či střední absolutní odchylkou předpovědi

$$\text{MADP} = \frac{1}{m} \sum_{t=n+1}^{n+m} \left| \hat{e}_t^{(p)} \right|. \quad (2.28)$$

Obdobně jako v předešlém případě, nižší hodnoty charakteristik (2.25), (2.26), (2.27) a (2.28) reflektují kvalitnější vztah modelu k budoucím hodnotám datového souboru.

Mezi charakteristiky, sloužící k porovnání různých modelů, řadíme Bayesovo informační kritérium (BIC) a Akaikeho informační kritérium (AIC). Obě charakteristiky jsou založené na věrohodnostní funkci (likelihood function), přitom výběr modelu spočívá v minimalizaci penalizační funkce. Tvary příslušných charakteristik jsou následující:

$$\text{BIC} = -2\ln(\hat{L}) + \ln(n) \cdot c, \quad (2.29)$$

$$\text{AIC} = -2\ln(\hat{L}) + K \cdot c, \quad (2.30)$$

kde \hat{L} představuje maximální hodnotu věrohodností funkce. K nejčastěji volíme ve tvaru $K = 2$, pokud dosadíme $K = \ln(n)$, obdržíme BIC. Celkový počet parametrů v modelu, zahrnující počet parametrů střední hodnoty a rozptylu, značíme symbolem c . V našem případě $c = p + 1$. Pro praktické výpočty se výraz $-2\ln(\hat{L})$ aproximuje výrazem $n \cdot \ln(\text{RSČ}/n)$.

2.4.2. Trendová složka

Trendová složka odráží dlouhodobé změny ČŘ. Může se jednat o dlouhodobý růst, či pokles. Model, který obsahuje pouze trendovou složku, se jeví jako velmi jednoduchý. Je vhodné jej využít pro průzkumovou analýzu ČŘ. Matematický model uvádíme v aditivním tvaru $y_t = Tr_t + \epsilon_t$. Nyní velmi stručně představíme základní typy trendů a jejich odhady a charakteristiky.

Konstantní trend ($Tr_t = \beta_0, t = 1, \dots, n, p = 1$)

Jedná se o trend, kdy data proložíme konstantní funkcí. Odhad parametru β_0 získáme vztahem $\hat{\beta}_0 = \bar{y}$, kde odvození parametru můžeme provést pomocí MNČ:

$$\min_{\beta_0} \sum_{t=1}^n (y_t - Tr_t)^2 = \min_{\beta_0} \sum_{t=1}^n (y_t - \beta_0)^2. \quad (2.31)$$

Derivací vztahu (2.31) dle parametru β_0 obdržíme normální rovnici

$$\begin{aligned} \frac{\partial}{\partial \beta_0} : 2 \sum_{t=1}^n (y_t - \beta_0)(-1) &= 0 \\ \sum_{t=1}^n y_t - n\beta_0 &= 0 \end{aligned} \quad (2.32)$$

a z rovnice (2.32) vyjádříme odhad β_0

$$\hat{\beta}_0 = \frac{\sum_{t=1}^n y_t}{n} = \bar{y}.$$

Předpověď pro tento trend je také konstanta $\hat{y}^{(p)} = \hat{\beta}_0$. Intervalovou předpověď získáme dle vzorců (2.12) a (2.13), kde $\hat{\sigma} = \sqrt{\frac{RS\check{C}}{n-1}} = \sqrt{\frac{\sum_{t=1}^n (y_t - \bar{y})^2}{n-1}}$ a datová matice X , resp. $X_{(p)}$ se skládá z jednotkového vektoru o rozměru $n \times 1$, resp. o rozměru $m \times 1$.

Lineární trend ($Tr_t = \beta_0 + \beta_1 t, t = 1, \dots, n, p = 2$)

Tento trend využíváme, když první diference $y_{t+1} - y_t$ jsou přibližně konstantní. Jedná se o proložení dat lineární funkcí $y_t = \beta_0 + \beta_1 t$. Odhad parametrů

dostaneme řešením soustavy normálních rovnic (dále NR). Nejprve provedeme minimalizaci výrazu ve smyslu MNČ:

$$\min_{\beta_0, \beta_1} \sum_{t=1}^n (y_t - Tr_t)^2 = \min_{\beta_0, \beta_1} \sum_{t=1}^n (y_t - \beta_0 - \beta_1 t)^2$$

a parciální derivací dle parametrů dostaneme soustavu NR:

$$\frac{\partial}{\partial \beta_0} : 2 \sum_{t=1}^n (y_t - \beta_0 - \beta_1 t)(-1) \doteq 0$$

$$\frac{\partial}{\partial \beta_1} : 2 \sum_{t=1}^n (y_t - \beta_0 - \beta_1 t)(-t) \doteq 0.$$

Soustavu NR upravíme do tvaru

$$\sum_{t=1}^n y_t - n\beta_0 - \beta_1 \sum_{t=1}^n t \doteq 0$$

$$\sum_{t=1}^n ty_t - \beta_0 \sum_{t=1}^n t - \beta_1 \sum_{t=1}^n t^2 \doteq 0.$$

Řešením této soustavy obdržíme odhady parametrů:

$$\hat{\beta}_1 = \frac{n \sum_{t=1}^n ty_t - \sum_{t=1}^n ty_t}{n \sum_{t=1}^n t^2 - (\sum_{t=1}^n t)^2} = \frac{n(\sum_{t=1}^n ty_t - \frac{1}{n} \sum_{t=1}^n ty_t)}{n(\sum_{t=1}^n t^2 - n(\frac{1}{n} \sum_{t=1}^n t)^2)} = \frac{\sum_{t=1}^n ty_t - \bar{t} \sum_{t=1}^n y_t}{\sum_{t=1}^n t^2 - n\bar{t}^2}, \text{ kde } \bar{t} = \frac{\sum_{t=1}^n t}{n},$$

$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{t}$, kde $\bar{y} = \frac{\sum_{t=1}^n y_t}{n}$. Odhad $\hat{\beta}_1$ jsme přitom spočítali pomocí Cramerova pravidla a odhad $\hat{\beta}_0$ jsme vyjádřili z první rovnice soustavy NR. Tyto odhady lze vypočítat také aplikací odhadu $\hat{\beta}$ dle (2.6), přičemž datové matice X a $X_{(p)}$ jsou ve tvaru:

$$X = \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ \vdots & \vdots \\ 1 & n \end{pmatrix}, \quad (2.33)$$

$$X_{(p)} = \begin{pmatrix} 1 & n+1 \\ 1 & n+2 \\ \vdots & \vdots \\ 1 & n+m \end{pmatrix}. \quad (2.34)$$

Předpovědi dostaneme dosazením do vztahu (2.11), neboli $\hat{y}_t^{(p)} = \hat{\beta}_0 + \hat{\beta}_1 t$, kde $t = n+1, \dots, n+m$. Intervalové předpovědi získáme pomocí vztahů (2.12) a (2.13).

Kvadratický trend ($Tr_t = \beta_0 + \beta_1 t + \beta_2 t^2$, $t = 1, \dots, n$, $p = 3$)

Tento trend je vhodný pro ČŘ, pro které jsou přibližně konstantní druhé diference $y_{t+2} - y_{t+1} + y_t$. Při použití tohoto trendu proložíme data parabolou.

Odhad parametrů bychom mohli opět provést vyřešením soustavy tří NR. Jelikož by byly samotné vzorce jednotlivých složek odhadů parametrů poněkud složité, efektivnějším způsobem je spočítat $\hat{\beta}$ dle (2.6), kde $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2)^T$. Datové matice X a $X_{(p)}$ uvažujeme ve tvaru

$$X = \begin{pmatrix} 1 & 1 & 1^2 \\ 1 & 2 & 2^2 \\ \vdots & \vdots & \vdots \\ 1 & n & n^2 \end{pmatrix}, \quad X_{(p)} = \begin{pmatrix} 1 & n+1 & (n+1)^2 \\ 1 & n+2 & (n+2)^2 \\ \vdots & \vdots & \vdots \\ 1 & n+m & (n+m)^2 \end{pmatrix}.$$

Bodové předpovědi obdržíme vztahem (2.11), tedy $\hat{y}_t^{(p)} = \hat{\beta}_0 + \hat{\beta}_1 t + \hat{\beta}_2 t^2$, kde $t = n+1, \dots, n+m$. Intervalové předpovědi získáme pomocí vztahů (2.12) a (2.13).

Exponenciální trend ($Tr_t = \alpha \beta^t$, kde $t = 1, \dots, n$, $\beta > 0$, $p = 2$)

Jedná se o trend, při kterém data proložíme exponenciální křivkou. Vyznačuje se tím, že podíl dvou sousedních diferencí má konstantní hodnotu β , tj. $\beta = \frac{Tr_{t+2} - Tr_{t+1}}{Tr_{t+1} - Tr_t}$. Kvůli odhadu parametrů musíme předpis trendu transformovat pomocí logaritmování na lineární trend

$$\ln Tr_t = \underbrace{\ln \alpha}_{a_0} + t \underbrace{\ln \beta}_{a_1},$$

kde odhad vektoru parametrů $a = (\ln \alpha, \ln \beta)^T$ vypočítáme pomocí MNČ dle vzorce (2.6). Odhady parametrů α a β dostaneme opětovným převedením na původní měřítko, tj. $\hat{\alpha} = e^{\hat{a}_0}$ a $\hat{\beta} = e^{\hat{a}_1}$. Datové matice X , resp. $X_{(p)}$ jsou opět ve tvaru (2.33), resp. (2.34). Logaritmované vyrovnané hodnoty spočítáme dle vztahu

$$\hat{y}^{(\ln)} = X\hat{a},$$

nebo zápisem po složkách

$$\hat{y}_t^{(\ln)} = \hat{a}_0 + \hat{a}_1 t, \quad t = 1, \dots, n, \quad (2.35)$$

přičemž symbol (\ln) zdůrazňuje spojitost s logaritmovanými vyrovnanými hodnotami. Vyrovnané hodnoty získáme opět aplikováním exponenciální funkce na logaritmované vyrovnané hodnoty

$$\hat{y} = e^{X\hat{a}}.$$

Bodové předpovědi získáme dosazením \hat{a} a $\hat{\beta}$ do trendu pro $t = n+1, \dots, n+m$, případně můžeme využít vektorový zápis $\hat{y}^{(p)} = e^{(X_{(p)}\hat{a})}$. Intervalové předpovědi pak obdržíme dle vzorců (2.12) a (2.13), kde meze předpovědního oboustranného intervalu jsou rovny

$$e^{\hat{y}_t^{(\ln)} \pm t_{(n-p), 1-\frac{\alpha}{2}} \hat{\sigma}^{(\ln)} \sqrt{1+(x_t^T (X^T X)^{-1} x_t)}}, \quad \text{kde } t = n+1, \dots, n+m \quad (2.36)$$

a meze předpovědního oboustranného simultánního intervalu vyjádříme vztahem

$$e^{\hat{y}_t^{(\ln)} \pm \hat{\sigma}^{(\ln)} \sqrt{m \cdot F_{(m, n-p), 1-\frac{\alpha}{2}} (1+x_t^T (X^T X)^{-1} x_t)}}, \quad \text{kde } t = n+1, \dots, n+m. \quad (2.37)$$

přičemž datové matice X a $X_{(p)}$ jsou stejné jako u lineárního trendu, $\hat{\sigma}^{(\ln)} = \sqrt{\frac{\hat{e}^{(\ln)T} \hat{e}^{(\ln)}}{n-2}}$ a $\hat{e}^{(\ln)} = \ln y - \hat{y}^{(\ln)}$. Můžeme také zvolit tzv. modifikovaný exponenciální trend, který má předpis $Tr_t = \gamma + \alpha \beta^t$, $\beta > 0$, $p = 3$, viz [5].

Logistický trend ($Tr_t = \frac{\gamma}{1+\alpha\beta^t}$, kde $t = 1, \dots, n$, $\beta > 0$, $\gamma > 0$, $p = 3$)

Jedná se o speciální typ S-křivky, která je symetrická kolem inflexního bodu $I = -\frac{\ln \alpha}{\ln \beta}$. Logistický trend můžeme chápat jako inverzi modifikovaného exponenciálního trendu. Logistického trendu využíváme, když se křivka prvních diferencí podobá křivce normální hustoty, nebo podíly $\frac{(Tr_{t+2})^{-1} - (Tr_{t+1})^{-1}}{(Tr_{t+1})^{-1} - (Tr_t)^{-1}}$ jsou přibližně konstantní. Abychom mohli odhadnout parametry, je potřeba provést drobné úpravy:

$$Tr_t = \frac{\gamma}{1 + \alpha\beta^t} \Big|^{-1}$$

$$\frac{1}{Tr_t} = \frac{1 + \alpha\beta^t}{\gamma}$$

$$\frac{1}{Tr_t} = \gamma^* + \alpha^*\beta^t \rightarrow \frac{1}{Tr_t} = \gamma^* + \alpha^*b^t.$$

Transformací původního předpisu dostaneme předpis pro lineární trend. Jelikož máme v modelu tři proměnné, je zapotřebí místo parametru β uvažovat pevnou konstantu $b \in (0, 1)$ takovou, aby při různých volbách konstanty b došlo k minimalizaci reziduálního součtu čtverců. Odhad parametrů a získáme aplikací vzorce (2.6), tedy $\hat{a} = (X^T X)^{-1} X^T \frac{1}{y}$, kde $\hat{a} = \begin{pmatrix} \hat{\gamma}^* \\ \hat{\alpha}^* \end{pmatrix}$, $\hat{\gamma} = \frac{1}{\hat{\gamma}^*}$ a $\hat{\alpha} = \hat{\alpha}^* \hat{\gamma}$.

Datové matice X a $X_{(p)}$ jsou ve tvaru

$$X = \begin{pmatrix} 1 & b^1 \\ 1 & b^2 \\ \vdots & \vdots \\ 1 & b^n \end{pmatrix}, \quad X_{(p)} = \begin{pmatrix} 1 & b^{n+1} \\ 1 & b^{n+2} \\ \vdots & \vdots \\ 1 & b^{n+m} \end{pmatrix}.$$

Vyrovnané hodnoty obdržíme dosazením odhadů $\hat{\alpha}$, $\hat{\gamma}$ a pevně zvolené konstanty b do trendové funkce, tj. $\hat{y}_t = \frac{\hat{\gamma}}{1+\hat{\alpha}b^t}$, pro čas $t = 1, \dots, n$. Obdobně postupujeme i u bodových předpovědích, jen pro čas $t = n+1, \dots, n+m$, případně $\hat{y}^{(p)}$ odpovídá převráceným hodnotám vektoru $X_{(p)}\hat{a}$. Za předpokladu normality převrácených hodnot y_t zkonstruujeme intervalových předpovědích dle vzorců (2.12) a (2.13), kde meze předpovědního oboustranného intervalu vyjádříme vztahem

$$\frac{1}{\hat{y}_t^{(\log)} \pm t_{(n-p), 1-\frac{\alpha}{2}} \widehat{\sigma}^{(\log)} \sqrt{1 + (x_t^T (X^T X)^{-1} x_t)}}, \quad \text{kde } t = n+1, \dots, n+m \quad (2.38)$$

a meze předpovědního oboustranného simultánního intervalu jsou rovny

$$\frac{1}{\hat{y}_t^{(\log)} \pm \hat{\sigma}^{(\log)} \sqrt{m \cdot F_{(m, n-p), 1-\frac{\alpha}{2}}(1 + x_t^T (X^T X)^{-1} x_t)}}, \text{ kde } t = n+1, \dots, n+m, \quad (2.39)$$

přičemž $\hat{y}_t^{(\log)}$ označuje logistické vyrovnané hodnoty, které spočítáme dle vztahu

$$\hat{y}_t^{(\log)} = \hat{\gamma}^* + \hat{\alpha}^* b^t, \text{ pro } t = 1, \dots, n,$$

symbol (\log) zdůrazňuje spojitost s logistickými vyrovnanými hodnotami, $\hat{\sigma}^{(\log)} = \sqrt{\frac{\hat{e}^{(\log)T} \hat{e}^{(\log)}}{n-2}}$ a $\hat{e}_t^{(\log)} = \frac{1}{y_t} - \hat{y}_t^{(\log)}$, $t = 1, \dots, n$.

Poznámka 2.1. Odhadu parametrů lze dosáhnout více způsoby. Můžeme vycházet z jiného řešení modifikovaného exponenciálního trendu, kdy data rozdělíme na třetiny. Dostaneme tři trendové rovnice, pro které odhadneme parametry α , β a γ . Další možností je využití tzv. principu diferenčních odhadů parametrů, kdy místo s původní řadou y_t pracujeme s řadou prvních diferencí $y_{t+1} - y_t$ (viz např. [5], str. 38).

Poslední trendovou funkcí, kterou uvedeme, je **Gompertzova křivka**. Jedná se také o S-křivku, jejíž předpis je ve tvaru $\ln T r_t = \gamma + \alpha \beta^t$, kde $t = 1, \dots, n$, $\beta > 0$. Odhad parametrů u Gompertzovy křivky provádíme obdobně jako u logistického trendu, jen místo převrácených hodnot řady y_t pracujeme s hodnotami $\ln y_t$. Inflexní bod Gompertzovy křivky je $I = -\frac{\ln \alpha}{\ln \beta}$.

Poznámka 2.2. Trend můžeme také popsat pomocí adaptivních metod, které si popíšeme v kapitole 2.4.4.

2.4.3. Periodická složka

Periodická složka se skládá ze součtu sezónní a cyklické složky. Sezónní složka má krátkodobý charakter (tj. výkyvy od trendu jsou kratší nebo přesně jeden rok). Oproti tomu cyklická složka je charakteristická pravidelnými výkyvy, které

jsou delší než jeden rok. Existuje celá řada přístupů, jak modelovat sezónní složku, ať už v multiplikatívním, tak v aditivním tvaru. Sezónnost můžeme popsat opět pomocí regresních modelů, přičemž musíme rozlišovat dva druhy sezónnosti, a to konstantní nebo proporcionální. Konstantní typ sezónnosti používáme, když amplituda ČŘ je konstantní v čase. Naproti tomu proporcionální typ sezónnosti je úměrný trendu, se kterým sezónnost buď roste, nebo klesá. Jelikož se v našich analyzovaných datových souborech nevyskytuje proporcionální typ sezónnosti, nebudeme se modely v multiplikatívním tvaru dále zabývat. Nyní představíme model s umělými proměnnými v aditivním tvaru, který z ČŘ eliminuje pouze sezónní složku.

Model s umělými proměnnými

Model s umělými proměnnými je charakterizován stálou (konstantní) sezónností vůči trendu. V praxi se projevuje stálými výkyvy od trendu mezi jednotlivými roky, čili sezónnost je v každém období jednotlivých roků konstantní. Model zapisujeme ve tvaru

$$y_{ij} = Tr_{ij} + S_j + E_{ij}, \quad (2.40)$$

kde $i = 1, \dots, I$ značí roky a $j = 1, \dots, r$ jsou jednotlivé sezóny v rámci roku, přičemž platí normalizační podmínka

$$\sum_{j=1}^r S_j = 0. \quad (2.41)$$

Místo S_{ij} stačí uvádět S_j , protože $S_{ij} = S_j$. Předpokládejme, že $n = I \cdot r$ a pro index t platí, že $t = (i - 1)r + j$, pak můžeme model (2.40) napsat ve tvaru

$$y_t = Tr_t + S_t + E_t, \quad t = 1, \dots, n. \quad (2.42)$$

Výhodou tohoto modelu je možnost kombinace různých trendových funkcí se sezónností.

Sezónní složku můžeme vyjádřit ve tvaru

$$S_t = \alpha_2 x_{2,t} + \alpha_3 x_{3,t} + \dots + \alpha_r x_{r,t}, \quad t = 1, \dots, n, \quad (2.43)$$

kde $x_{2,t}, \dots, x_{r,t}$ jsou indikátorové proměnné, které jsou definované takto:

$$x_{j,t} = \begin{cases} 1, & \text{čas } t \text{ odpovídá } j\text{-té sezóně v roce} \\ 0, & \text{jinak} \end{cases}$$

a $\alpha_2, \dots, \alpha_r$ jsou regresní parametry příslušející umělým proměnným. Všimněme si, že v modelu nemůžeme uvažovat α_1 , jelikož by sloupce matice plánu byly lineárně závislé. To odpovídá přesné kolinearitě v datech, která má za následek, že nelze spočítat odhad MNČ. První sezóna je označována jako referenční. Matici plánu obdržíme sloučením sloupců trendové matice plánu a matice plánu umělých proměnných. Např. pokud bychom uvažovali kvadratický trend a čtvrtletní model sezónnosti ($m = 4$) pro čtvrtletní data, matice plánů X a $X_{(p)}$ pro jeden následující rok by byly ve tvaru:

$$X = \begin{pmatrix} 1 & 1 & 1^2 & 0 & 0 & 0 \\ 1 & 2 & 2^2 & 1 & 0 & 0 \\ 1 & 3 & 3^2 & 0 & 1 & 0 \\ 1 & 4 & 4^2 & 0 & 0 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & n & n^2 & 0 & 0 & 1 \end{pmatrix}, X_{(p)} = \begin{pmatrix} 1 & n+1 & (n+1)^2 & 0 & 0 & 0 \\ 1 & n+2 & (n+2)^2 & 1 & 0 & 0 \\ 1 & n+3 & (n+3)^2 & 0 & 1 & 0 \\ 1 & n+4 & (n+4)^2 & 0 & 0 & 1 \end{pmatrix}.$$

Odhad vektoru parametrů a spočítáme pomocí vzorce (2.6), přičemž $\hat{a} = (\hat{\beta}_0, \dots, \hat{\beta}_k, \hat{\alpha}_2, \dots, \hat{\alpha}_r)^T$. Poté je nutné odhady $\hat{\alpha}_j$ centrovat, tedy od jednotlivých odhadů odečteme průměr $\bar{\alpha} = \frac{\hat{\alpha}_2 + \dots + \hat{\alpha}_r}{r}$. Zároveň dostaneme r sezónních faktorů $\hat{S}_j = \hat{\alpha}_j - \bar{\alpha}$, $j = 1, \dots, r$, kde $\hat{\alpha}_1 = 0$. Dále nesmíme opomenout upravit absolutní člen do tvaru: $\hat{\beta}_0^* = \hat{\beta}_0 + \bar{\alpha}$. Model (2.42) můžeme rozepsat díky vztahu (2.43) do tvaru

$$y_t = Tr_t + \alpha_2 x_{2,t} + \alpha_3 x_{3,t} + \dots + \alpha_r x_{r,t} + \epsilon_t, \quad t = 1, \dots, n. \quad (2.44)$$

Vyrovnané hodnoty \hat{y}_t získáme dosazením odhadů do modelu (2.44) pro čas $t = 1, \dots, n$. V případě, že bychom chtěli vypočítat bodové předpovědi, postupujeme obdobně jako u vyrovnaných hodnot, jen pro čas $t = n+1, \dots, n+m$.

Intervalové předpovědi obdržíme pomocí vztahů (2.12) a (2.13), přičemž $\hat{\sigma}^2$ vypočítáme dle (2.16) a $p = k + r - 1$.

Model skrytých period

Předchozí model s umělými proměnnými výborně popisoval jak trend, tak sezónnost, nicméně se v něm nepromítanou výraznější odchylky od trendu, které jsou delší než jeden rok. Těmto odchylkám říkáme cykly. Pro modelování jak sezónnosti, tak i cyklu poslouží přístup spektrální analýzy, a to model skrytých period (dále MSP). MSP namodeluje ČŘ prostřednictvím goniometrických křivek s různými frekvencemi a amplitudami. Tento model nerozlišuje cyklus nebo sezónnost, ale do výsledného modelu zahrnuje pouze významné frekvence, jež periodicitu namodelují. Nejprve však uvedeme základní pojmy spektrální analýzy.

Frekvence jsou periodické funkce času t . Frekvence se udávají v [Hz]³ za časovou jednotku, což je časový interval mezi dvěma pozorováními ČŘ. Např. u T -periodické funkce $\sin(\frac{2\pi t}{T})$ nebo $\cos(\frac{2\pi t}{T})$ by jeden cyklus proběhl za dobu časové jednotky T . Frekvence tak udává počet cyklů za časovou jednotku. Platí, že čím je frekvence větší, tím častěji se v průběhu ČŘ střídají jednotlivé cykly. Abychom rozlišili pojem frekvence z fyzikálního hlediska, budeme ω označovat jako úhlovou frekvenci, kde $\omega = 2\pi \cdot \text{frekvence} = \frac{2\pi \cdot \text{počet cyklů}}{\text{časová jednotka}}$, jež je udávána v [rad]⁴. Jelikož uvažujeme ČŘ na intervalu $\langle 0, n \rangle$, tak $\omega = \frac{2\pi j}{n}, j = 1, \dots, J$, kde

$$J = \begin{cases} \frac{n}{2} & n \text{ je sudé} \\ \frac{n-1}{2} & n \text{ je liché} . \end{cases}$$

Perioda (cyklus) představuje převrácenou hodnotu frekvence, tedy $\text{perioda} = \frac{1}{\text{frekvence}}$. Jinými slovy, perioda je doba, za kterou proběhne právě jedna vlna.

MSP je ve tvaru

$$y_t = \alpha_0 + \sum_{j=1}^J \alpha_j \sin(\omega_j t) + \sum_{j=1}^J \beta_j \cos(\omega_j t) + \epsilon_t, t = 1, \dots, n. \quad (2.45)$$

Aby nebyl model (2.45) příliš složitý, budeme v modelu uvažovat pouze významné frekvence, které určíme pomocí periodogramu.

³[Hz] značí hertz.

⁴[rad] představuje radiány.

Periodogram ČŘ se definuje jako funkce proměnné ω ve tvaru

$$I(\omega) = \frac{1}{4\pi}(a^2(\omega) + b^2(\omega)), \quad -\pi \leq \omega \leq \pi, \quad (2.46)$$

kde

$$a(\omega) = \sqrt{\frac{2}{n}} \sum_{t=1}^n y_t \sin(\omega t), \quad b(\omega) = \sqrt{\frac{2}{n}} \sum_{t=1}^n y_t \cos(\omega t). \quad (2.47)$$

Jednotlivé frekvence otestujeme Fisherovým testem periodicity, který testuje nulovou hypotézu ve tvaru $H_0 : y_t = \alpha_0 + \epsilon_t$, což znamená, že ČŘ neobsahuje žádnou periodickou složku, jinými slovy ČŘ je úměrná $\epsilon \sim WN(0, \sigma^2 I)$. V souladu s H_0 by úhlová frekvence ω_j neměla být významná. Alternativní hypotéza H_1 tvrdí, že alespoň jedna úhlová frekvence ω_j významná je, tj. $\exists j \in 1, \dots, J : \alpha_j \neq 0 \vee \beta_j \neq 0$.

Postup nalezení významných frekvencí spočívá v napočítání hodnot $I(\omega_j)$, kde $j = 1, \dots, J$, které uspořádáme dle velikosti, tj.

$$I(\omega_1) \geq I(\omega_2) \geq \dots \geq I(\omega_J).$$

Dále postupně počítáme normovanou testovou statistiku pro uspořádané hodnoty $I(\omega_j)$, $j = 1, \dots, J$, tedy

$$Y_{(j)} = \frac{I(\omega_j)}{\sum_{i=1}^J I(\omega_i)}, \quad \text{kde } j = 1, \dots, J. \quad (2.48)$$

Pokud hodnota $Y_1 > g_{p,1-\alpha}^F$, pak zamítáme H_0 na hladině významnosti α (jedná se o jednostranný test), kde $g_{p,1-\alpha}^F$ je kritická hodnota Fisherova testu, již lze najít ve speciálních tabulkách [15]. Zamítnutí H_0 svědčí o tom, že ω_1 je významná frekvence. Dále pokračujeme pro $j = 2, \dots, J$, dokud H_0 nelze zamítnout. V tom případě test končí. Při výpočtu dalších hodnot testové statistiky $Y_{(j)}$ nesmíme opomenout vynechat vždy příslušnou významnou frekvenci z daného vzorce (2.48). Dostaneme tak V významných frekvencí.

Výsledný model je v podobném tvaru jako (2.45), ale bereme v úvahu pouze významné frekvence ω_j , $j = 1, \dots, V$. V případě, že obdržíme více významných

frekvencí, je vhodné zvážit, zda by nám v modelu nestačily první dvě významné frekvence. Všimněme si, že v modelu máme absolutní člen α_0 , který představuje konstantní trend. MSP můžeme vhodně modifikovat nahrazením konstantního trendu jiným trendem (pro předpovědi je vhodný např. MSP s lineárním trendem). U složitějších trendových funkcí se datový soubor nejdříve očistí od trendu a následně se na rezidua aplikuje MSP s konstantním trendem.

Jelikož je model (2.45) lineární v parametrech, můžeme příslušné odhady parametrů spočítat dle MNČ s využitím vzorce (2.6). Dostaneme odhad parametrů $\hat{\gamma} = (\hat{\alpha}_0, \hat{\alpha}_1, \dots, \hat{\alpha}_V, \hat{\beta}_1, \dots, \hat{\beta}_V)^T$. Matice plánu X je ve tvaru

$$X = \begin{pmatrix} 1 & \sin(\omega_1 \cdot 1) & \dots & \sin(\omega_V \cdot 1) & \cos(\omega_1 \cdot 1) & \dots & \cos(\omega_V \cdot 1) \\ 1 & \sin(\omega_1 \cdot 2) & \dots & \sin(\omega_V \cdot 2) & \cos(\omega_1 \cdot 2) & \dots & \cos(\omega_V \cdot 2) \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 1 & \sin(\omega_1 \cdot n) & \dots & \sin(\omega_V \cdot n) & \cos(\omega_1 \cdot n) & \dots & \cos(\omega_V \cdot n) \end{pmatrix}. \quad (2.49)$$

Matici plánu pro předpovědi $X_{(p)}$ dostaneme obdobně jako matici X , ale pro čas $t = n + 1, \dots, n + m$. Pokud bychom uvažovali složitější trend, doplníme matici plánů daného trendu do matice (2.49) místo prvního sloupce, jenž v tomto případě přísluší konstantnímu trendu.

Vyrovnané hodnoty obdržíme dosazením odhadu $\hat{\gamma}$ do modelu (2.45) pro úhlové frekvence $\omega_j, j = 1, \dots, V$. Bodové předpovědi vypočítáme obdobně jako vyrovnané hodnoty, jen pro čas $t = n + 1, \dots, n + m$. Intervalové předpovědi získáme aplikací vzorců (2.12) a (2.13), přičemž $\hat{\sigma}^2$ vypočítáme dle (2.16). Jelikož nyní uvažujeme konstantní trend, tak $p = 1 + 2V$.

2.4.4. Adaptivní metody

V předchozím textu jsme model založili pouze na historických datech. Abychom získali aktuální model z dat, museli jsme napočítat parametry modelu. Adaptivní metody se naopak přizpůsobují přidáním pozorování, čímž dochází k aktualizaci modelu. Flexibilita modelu je tak založená na lokální eliminaci trendu. Mezi nejznámější metody řadíme klouzavé průměry, exponenciální

vyrovnávání a Holtovu-Wintersovu metodu. Metoda klouzavých průměrů je charakteristická subjektivní volbou délky okna. V rámci této práce si proto představíme poslední dvě z uvedených metod, jenž tento problém překonávají. V této podkapitole vycházím z [5, 9, 10, 11, 12].

Exponenciální vyrovnávání

Obecně budeme uvažovat model ve tvaru

$$y_{t-k} = Tr_{t-k} + e_{t-k}, k = 0, \dots, t-1, \quad (2.50)$$

přičemž pro odhad parametrů využijeme váženou MNČ, kde se váhy jednotlivých čtverců ze součtu snižují v závislosti na „stáří“ pozorování, které označujeme symbolem k . Nejnovější pozorování dostane největší váhu, naopak u starších pozorování váhu snižujeme. Váha je charakterizována vyrovnávací konstantou α , přičemž $\alpha \in (0, 1)$. Čas t představuje aktuální okamžik.

Vážená MNČ je ve tvaru

$$\min \sum_{k=0}^{t-1} \alpha^k (y_{t-k} - Tr_{t-k})^2, \quad (2.51)$$

kde α jsou exponenciálně klesající váhy.

a) Jednoduché exponenciální vyrovnání ($Tr_t = \beta_0$, model nezávisí na k)

Tento model je založen na lokálním vyhlazení ČŘ konstantním trendem. Vztah (2.53) odvodíme z výrazu (2.51) dosazením β_0 do trendové funkce

$$\min \sum_{k=0}^{t-1} \alpha^k (y_{t-k} - \beta_0)^2,$$

Derivací minimalizační funkce dle β_0 obdržíme normální rovnici

$$\frac{\partial}{\partial \beta_0} : -2 \sum_{k=0}^{t-1} \alpha^k (y_{t-k} - \beta_0) \doteq 0,$$

kterou upravíme do následující rovnice

$$\beta_0 \sum_{k=0}^{t-1} \alpha^k = \sum_{k=0}^{t-1} \alpha^k y_{t-k}. \quad (2.52)$$

Dále předpokládejme, že $\sum_{k=0}^{t-1} \alpha^k \doteq 0$, $\sum_{k=0}^{\infty} \alpha^k = \frac{1}{1-\alpha}$, potom rovnici (2.52) přepíšeme do tvaru

$$\beta_0 \frac{1}{1-\alpha} = \sum_{k=0}^{t-1} \alpha^k y_{t-k},$$

ze kterého již vyjádříme odhad parametru β_0 v čase t

$$\widehat{\beta}_0(t) = (1-\alpha) \sum_{k=0}^{t-1} \alpha^k y_{t-k}. \quad (2.53)$$

Výraz 2.53 představuje jednoduchou vyrovnávací statistiku S_t a zároveň vyrovnanou hodnotu v čase t .

Vyrovnané hodnoty spočítáme také na základě aktualizací vzorce jako vážený průměr aktuálního pozorování a předchozí vyrovnávací hodnoty

$$\widehat{y}_t = (1-\alpha) y_t + \alpha \widehat{y}_{t-1}, \quad t = 1, \dots, n. \quad (2.54)$$

Odvození vyrovnané hodnoty v čase t dle vzorce (2.54) je následující:

$$\begin{aligned} \widehat{y}_t &= (1-\alpha) \sum_{k=0}^{t-1} \alpha^k y_{t-k} \\ &= (1-\alpha) y_t + (1-\alpha) \sum_{k=1}^{t-1} \alpha^k y_{t-k} \\ &= (1-\alpha) y_t + \alpha (1-\alpha) \sum_{k=1}^{t-1} \alpha^{k-1} y_{t-1-(k-1)} \\ &\stackrel{j=k-1}{=} (1-\alpha) y_t + \alpha (1-\alpha) \underbrace{\sum_{j=0}^{t-2} \alpha^j y_{t-1-j}}_{\widehat{y}_{t-1}} \\ &= (1-\alpha) y_t + \alpha \widehat{y}_{t-1}. \end{aligned}$$

Pro stanovení vhodné vyrovnávací konstanty α použijeme dvoufázový postup. Zároveň abychom mohli aktualizací vzorec (2.54) použít, je nutné vhodně zvolit hodnotu \hat{y}_0 . V první fázi zvolíme $\hat{y}_0 = \frac{1}{6} \sum_{t=1}^6 y_t$ nebo $\hat{y}_0 = \frac{1}{n/2} \sum_{t=1}^{n/2} y_t$ a vypočítáme na vyrovnané hodnoty a hodnoty reziduálního součtu čtverců pro různé hodnoty α . Optimální volba α je taková, jež vede k minimalizaci RSC. V druhé fázi již počítáme s optimálním α , přičemž počáteční hodnota je průměr z celé časové řady, tedy $\hat{y}_0 = \frac{1}{n} \sum_{t=1}^n y_t$.

Bodová předpověď nezávisí na budoucím vývoji ČŘ, tedy $\hat{y}_t^{(p)} = \hat{y}_n$, $t = n + 1, \dots, n + m$. Meze $(1 - \alpha) \cdot 100\%$ předpovědního oboustranného intervalu zkonstruujeme dle [9] na základě vztahu

$$\hat{y}_t^{(p)} \pm u_{1-\alpha/2} \cdot \text{RMSE} \cdot d, \quad t = n + 1, \dots, n + m, \quad (2.55)$$

kde $u_{1-\alpha/2}$ je kvantil normovaného normálního rozdělení $N(0,1)$, RMSE představuje odmocninu ze střední čtvercové chyby dle vztahu (2.23) a $d = 1,25$. Meze oboustranného simultánního předpovědního intervalu jsou rovny

$$\hat{y}_t^{(p)} \pm \text{RMSE} \cdot \sqrt{mF_{(m,n-m),1-\frac{\alpha}{2}}} \cdot d, \quad t = n + 1, \dots, n + m. \quad (2.56)$$

Poznámka 2.3. Dle Cipry [5], lze ve vztazích (2.55) a (2.56) použít místo RMSE hodnotu MAD.

b) Dvojité exponenciální vyrovnání ($Tr_t = \beta_0 - \beta_1 k$, $t = 1, \dots, n - 1$)

Na rozdíl od jednoduchého exponenciálního vyrovnávání, se model v tomto případě opírá o lokální vyhlazení ČŘ lineárním trendem. Pro výpočet odhadů parametrů, budeme derivovat výraz (2.51) dle parametrů β_0 , β_1 , čímž dostaneme i vyrovnávací statistiky.

$$\min \sum_{k=0}^{t-1} \alpha^k (y_{t-k} - \beta_0 - \beta_1 k)^2,$$

$$\frac{\partial}{\partial \beta_0} : -2 \sum_{k=0}^{t-1} \alpha^k (y_{t-k} - \beta_0 - \beta_1 k) \doteq 0$$

$$\frac{\partial}{\partial \beta_1} : 2 \sum_{k=0}^{t-1} \alpha^k k (y_{t-k} - \beta_0 - \beta_1 k) \doteq 0.$$

Obdržíme tak soustavu NR:

$$\sum_{k=0}^{t-1} \alpha^k y_{t-k} = \beta_0 \sum_{k=0}^{t-1} \alpha^k - \beta_1 \sum_{k=0}^{t-1} \alpha^k k$$

$$\sum_{k=0}^{t-1} \alpha^k k y_{t-k} = \beta_0 \sum_{k=0}^{t-1} \alpha^k k - \beta_1 \sum_{k=0}^{t-1} \alpha^k k^2.$$

Předpokládejme, že

$$\text{a) } \sum_{k=1}^{t-1} \alpha^{k-1} k \doteq \sum_{k=1}^{\infty} \alpha^{k-1} k \xrightarrow{/\cdot \alpha} \sum_{k=0}^{\infty} \alpha^k k = \frac{\alpha}{(1-\alpha)^2},$$

$$\text{b) } \sum_{k=1}^{t-1} \alpha^{k-1} k^2 \doteq \sum_{k=1}^{\infty} \alpha^{k-1} k^2 = \frac{(1-\alpha)^2 + \alpha 2(1-\alpha)}{(1-\alpha)^4} = \frac{(1-\alpha) \cdot (1-\alpha + 2\alpha)}{(1-\alpha)^4} = \frac{1+\alpha}{(1-\alpha)^3},$$

z čehož vyplývá, že $\sum_{k=0}^{\infty} \alpha^k k^2 = \frac{\alpha \cdot (1+\alpha)}{(1-\alpha)^3}$.

Následně upravíme soustavu NR:

$$\sum_{k=0}^{t-1} \alpha^k y_{t-k} = \beta_0 \frac{1}{1-\alpha} - \beta_1 \frac{\alpha}{(1-\alpha)^2} \mid \cdot (1-\alpha)$$

$$(1-\alpha) \sum_{k=0}^{t-1} \alpha^k k y_{t-k} = \beta_0 \frac{\alpha}{(1-\alpha)^2} - \beta_1 \frac{\alpha(1+\alpha)}{(1-\alpha)^3} \mid \cdot (1-\alpha)^2,$$

kterou dále prepíšeme do následujících tvarů

$$(1-\alpha) \sum_{k=0}^{t-1} \alpha^k y_{t-k} = \beta_0 - \beta_1 \frac{\alpha}{(1-\alpha)}$$

$$(1-\alpha)^2 \sum_{k=0}^{t-1} \alpha^k k y_{t-k} = \beta_0 \alpha - \beta_1 \frac{\alpha(1+\alpha)}{1-\alpha}$$

Levá strana první rovnice soustavy NR představuje jednoduchou vyrovnávací statistiku S_t a můžeme ji zapsat ve tvaru

$$S_t = \beta_0 - \beta_1 \frac{\alpha}{(1 - \alpha)}. \quad (2.57)$$

Pro další účely zadefinujme dvojitou vyrovnávací statistiku

$$S_t^{(2)} = (1 - \alpha) \sum_{k=0}^{t-1} \alpha^k S_{t-k}. \quad (2.58)$$

Dále platí

$$\text{a) } S_t = (1 - \alpha) y_t + \alpha S_{t-1}$$

$$\text{b) } S_t^{(2)} = (1 - \alpha) S_t + \alpha S_{t-1}^{(2)} \rightarrow \alpha S_{t-1}^{(2)} = S_t^{(2)} - (1 - \alpha) S_t$$

$$\text{c) } \alpha S_{t-1}^{(2)} = (1 - \alpha)^2 \sum_{k=0}^{t-1} \alpha^k k y_{t-k}.$$

Přičemž vztahy a), b) odvodíme analogicky jako (2.54). Vztah c) odvodíme s využitím definice (2.58) následovně:

$$\begin{aligned} \alpha S_{t-1}^{(2)} &= \alpha (1 - \alpha) \sum_{k=0}^{t-1} \alpha^k S_{t-k} = \alpha (1 - \alpha) \sum_{k=0}^{t-2} \alpha^k S_{t-1-k} \\ &= \alpha (1 - \alpha) \sum_{k=0}^{t-2} \alpha^k (1 - \alpha) \sum_{j=0}^{t-1-1-k} \alpha^j y_{t-1-k-j} \\ &= \alpha (1 - \alpha)^2 \sum_{k=0}^{t-2} \sum_{j=0}^{t-2-k} \alpha^{k+j} y_{t-1-(k+j)} \\ &\stackrel{l=k+j}{=} \alpha (1 - \alpha)^2 \sum_{l=0}^{t-2} (l+1) \alpha^l y_{t-1-l} = (1 - \alpha)^2 \sum_{l=0}^{t-2} (l+1) \alpha^{l+1} y_{t-(l+1)} \\ &\stackrel{k=l+1}{=} (1 - \alpha)^2 \sum_{k=1}^{t-1} k \alpha^k y_{t-k} = (1 - \alpha)^2 \sum_{k=0}^{t-1} k \alpha^k y_{t-k} \end{aligned} \quad (2.59)$$

To nám umožní zapsat druhou rovnici soustavy NR pomocí vyrovnávacích statistik. Dostáváme tak

$$\alpha S_{t-1}^{(2)} = \beta_0 \alpha - \beta_1 \frac{\alpha(1+\alpha)}{1-\alpha}$$

$$S_t^{(2)} - (1-\alpha) S_t = \beta_0 \alpha - \beta_1 \frac{\alpha(1+\alpha)}{1-\alpha}. \quad (2.60)$$

Když dosadíme rovnici (2.57) do rovnice (2.60), můžeme $S_t^{(2)}$ vyjádřit ve tvaru:

$$\begin{aligned} S_t^{(2)} &= \beta_0 \alpha - \beta_1 \frac{\alpha(1+\alpha)}{1-\alpha} + (1-\alpha) S_t \\ &= \beta_0 \alpha - \beta_1 \frac{\alpha(1+\alpha)}{1-\alpha} + (1-\alpha) \left[\beta_0 - \beta_1 \left(\frac{1}{1-\alpha} \right) \right] \\ &= \beta_0 \alpha + (1-\alpha) \beta_0 - \beta_1 \frac{\alpha(1+\alpha)}{1-\alpha} - \beta_1 \alpha \\ &= \beta_0 (\alpha + 1 - \alpha) - \beta_1 \left(\frac{\alpha(1+\alpha)}{1-\alpha} + \alpha \right) \\ &= \beta_0 - \beta_1 \frac{2\alpha}{1-\alpha}. \end{aligned} \quad (2.61)$$

Odhady parametrů β_0 a β_1 vyjádříme z rovnic (2.57) a (2.60), čili

$$\widehat{\beta}_0(t) = 2S_t - S_t^{(2)}, \quad t = 1, \dots, n, \quad (2.62)$$

$$\widehat{\beta}_1(t) = \frac{1-\alpha}{\alpha} S_t - S_t^{(2)}, \quad t = 1, \dots, n. \quad (2.63)$$

Vyrovnané hodnoty obdržíme dosazením získaných odhadů (2.62), (2.63) do modelu, tedy $\widehat{y}_{t-k} = \widehat{\beta}_0(t) - \widehat{\beta}_1(t)k$, kde $k = 0, \dots, t-1$.

Poznámka 2.4. Pro aktuální čas $k = 0$ se vyrovnaná hodnota rovná odhadu $\widehat{\beta}_0(t)$, tj. $\widehat{y}_t = 2S_t - S_t^{(2)}$, pro $t = 1, \dots, n$.

Pro výpočet vyrovnaných hodnot je vhodné využít rekurentní vztahy vyrovnávacích statistik, přičemž startovací hodnoty $S_0, S_0^{(2)}$ vypočítáme tímto způsobem:

$$S_0 = \widehat{\beta}_0^0 - \widehat{\beta}_1^0 \frac{\alpha}{1-\alpha},$$

$$S_0^{(2)} = \hat{\beta}_0^0 - \hat{\beta}_1^0 \frac{\alpha^2}{1 - \alpha},$$

kde $\hat{\beta}_0^0$ a $\hat{\beta}_1^0$ jsou MNC odhady lineárního trendu z celé ČR.

Bodové předpovědi v čase n obdržíme dle vztahu

$$\hat{y}_{n+k}^{(p)} = \hat{\beta}_0(n) - \hat{\beta}_1(n)k, \text{ kde } k = 1, \dots, m. \quad (2.64)$$

$(1 - \alpha) \cdot 100\%$ oboustranný předpovědní interval je dán předpisem (2.55), přičemž hodnotu d nahradíme hodnotou d_h , $h = 1, \dots, m$, kde

$$d_h = 1.25 \frac{1 + \frac{1-\alpha}{(1+\alpha)^3} [(1 + 4\alpha + 5\alpha^2) + 2(1 - \alpha)(1 + 3\alpha)h + 2(1 - \alpha)^2 h^2]}{1 + \frac{1-\alpha}{(1+\alpha)^3} [(1 + 4\alpha + 5\alpha^2) + 2(1 - \alpha)(1 + 3\alpha) + 2(1 - \alpha)^2]}.$$

Obdobně postupujeme i v případě oboustranného předpovědního simultánního intervalu dle vztahu (2.56), ve kterém hodnotu d opět nahradíme hodnotou d_h .

Nicméně, musíme dále uvést také metody a způsob volby optimální vyrovnávací konstanty α . Opět využíváme dvoufázový postup, kde v první části napočítáme startovací hodnoty vyrovnávacích statistik z prvních šesti nebo z první poloviny pozorování. Následně počítáme hodnoty RSC pro různé hodnoty α . Optimální α je opět takové, které vede k minimalizaci RSC. Dále postup opakujeme již s optimálním α , kde startovací hodnoty vyrovnávacích statistik počítáme již z celé ČR.

Další možností je využít optimální α z jednoduchého exponenciálního vyrovnávání, nebo výraz $\hat{\alpha} = \sqrt{\frac{M-1}{M+1}}$, kde M představuje délku okna jednoduchého klouzavého průměru. Problém však nastává v subjektivní volbě M , které se chceme spíše vyvarovat.

Poznámka 2.5. Pro případ, že bychom lokální část ČR chtěli vyrovnat kvadratickým trendem, můžeme využít trojitě exponenciální vyrovnávání.

Aditivní Holtova-Wintersova metoda

Tento model je založen na kombinaci modelu s umělými proměnnými a exponenciálního vyrovnávání. Výhoda modelu tkví v zahrnutí sezónní složky. Prakticky tak jde o zobecnění exponenciálního vyrovnávání. V případě, že by sezónní složka nebyla přibližně konstantní, je vhodnější využít multiplikativní obdobu této metody (viz [5], str. 88), přičemž celkem je možné narazit na osm verzí této metody, viz [10]. Model uvažujeme ve tvaru

$$y_t = Tr_t + S_t + E_t, \text{ kde } t = 1, \dots, n, \quad (2.65)$$

přičemž Tr_t odpovídá trendu, S_t představuje sezónní složku. Abychom navázali na dvojité exponenciální vyrovnávání, předpokládáme, že trend se skládá z úrovně trendu L_t a směrnice trendu B_t . Počet sezón v jednom roce je r . Nyní představíme specifické formule pro výpočet jednotlivých složek.

Úroveň trendu v čase t je vážený průměr aktuálního pozorování očištěného od poslední známé odpovídající sezóny a součtu odhadu úrovně a směrnice trendu v čase $t - 1$:

$$L_t = \alpha(y_t - S_{t-r}) + (1 - \alpha)(L_{t-1} + B_{t-1}). \quad (2.66)$$

Směrnice trendu je vyjádřena váženým průměrem rozdílu úrovní trendu v čase t , $t - 1$ ⁵ a odhadu směrnice trendu v čase $t - 1$:

$$B_t = \beta(L_t - L_{t-1}) + (1 - \beta)B_{t-1}. \quad (2.67)$$

Sezónní složku popisujeme váženým průměrem aktuálního pozorování očištěného od aktuální hodnoty úrovně trendu a hodnoty sezónní složky odpovídající dané sezóně z předchozího roku:

$$S_t = \gamma(y_t - L_t) + (1 - \gamma)S_{t-r}. \quad (2.68)$$

Holtova-Wintersova metoda je založena na třech vyrovnávacích parametrech, kde α přísluší úrovni trendu, β směrnici trendu a γ sezónnosti.

Vyrovnané hodnoty obdržíme:

⁵Výraz $L_t - L_{t-1}$ představuje naivní odhad směrnice trendu v čase t .

a) součtem aktuálních hodnot úrovně a sezónnosti

$$\hat{y}_t = L_t + S_t, t = , \dots, n. \quad (2.69)$$

b) součtem hodnot úrovně, směrnici trendu a sezónnosti v čase $t - 1$

$$\hat{y}_t = L_{t-1} + B_{t-1} + S_{t-r}, t = , \dots, n. \quad (2.70)$$

Vyvstává však opět otázka zvolení optimálních hodnot vyrovnávacích parametrů. Jako u exponenciálního vyrovnávání, optimální variantou se jeví hodnoty parametrů, jež minimalizují RSČ. V případě, že potřebujeme rychle nastavit určité hodnoty, Ch. Chatfield a M. Yar [12] doporučují mimo jiné nastavit hodnoty na $\alpha = 0,3$, $\beta = 0,1$, $\gamma = 0,3$. Z jejich zkušenosti hodnoty parametrů α a γ často překračují hodnotu 0,3, naopak hodnota parametru β je často menší nebo rovna 0,1 a může být dokonce nulová. Nicméně, dle empirické studie [11], inicializační hodnoty těchto parametrů mají špatný efekt na přesnost predikcí. Abychom dostali kvalitní predikce, měli bychom vyrovnávací parametry zvolit na základě MSEP, nebo vybrat takové parametry, jež vedou k lokální minimalizaci nějaké z funkcí chyb předpovědi (např. MADP, MSEP).

Inicializační hodnoty formulí L_0 , B_0 , S_p , $p = 1, \dots, r$, můžeme nastavit dle odhadu $\hat{\alpha}$ z metody s umělými proměnnými (opět předpokládáme, že první sezóna je referenční, tj. $S_1 = 0$) nebo prostřednictvím vztahů $L_0 = \frac{1}{r} \sum_{t=1}^r y_t$, $B_0 = \frac{1}{r^2} \sum_{t=1}^r (y_{t+1} - y_t)$ a $S_p = y_p - L_p$ pro, $p = 1, \dots, r$.

Formule (2.66), (2.67), (2.68) a vyrovnané hodnoty si můžeme postupně zepsat do tzv. chybového tvaru prostřednictvím reziduí, jež vypočteme vztahem (2.14)

$$\hat{e}_t = y_t - L_{t-1} - B_{t-1} - S_{t-1}, t = , \dots, n. \quad (2.71)$$

Dále tak dostaneme upravené rekurentní vztahy pro L_t , B_t a S_t :

$$\begin{aligned} L_t &= \alpha(y_t - S_{t-r}) + (1 - \alpha)(L_{t-1} + B_{t-1}) \\ &= \alpha(e_t + L_{t-1} + B_{t-1}) + (1 - \alpha)(L_{t-1} + B_{t-1}) \\ &= \alpha e_t + \alpha L_{t-1} + \alpha B_{t-1} + (1 - \alpha)L_{t-1} + (1 - \alpha)B_{t-1} \\ &= \alpha e_t + L_{t-1} + B_{t-1}. \end{aligned}$$

S využitím předchozího vztahu si vyjádříme směrnicí trendu

$$\begin{aligned}
 B_t &= \beta(L_t - L_{t-1}) + (1 - \beta) B_{t-1} \\
 &= \beta(\alpha e_t + B_{t-1}) + (1 - \beta) B_{t-1} \\
 &= \alpha\beta e_t + \beta B_{t-1} + (1 - \beta) B_{t-1} \\
 &= \alpha\beta e_t + B_{t-1}.
 \end{aligned}$$

Pro odvození sezónní složky opět využijeme upravený vztah L_t

$$\begin{aligned}
 S_t &= \gamma(y_t - L_t) + (1 - \gamma) S_{t-r} \\
 &= \gamma(y_t - L_{t-1} - B_{t-1} - \alpha e_t) + (1 - \gamma) S_{t-r} \\
 &= \gamma(e_t + B_{t-1} + L_{t-1} + S_{t-r} - L_{t-1} - B_{t-1} - \alpha e_t) + (1 - \gamma) S_{t-r} \\
 &= (1 - \alpha) \gamma e_t + \gamma S_{t-r} + (1 - \gamma) S_{t-r} \\
 &= (1 - \alpha) \gamma e_t + S_{t-r}.
 \end{aligned}$$

Bodové předpovědi v čase $t = n + 1, \dots, n + m$ obdržíme vztahem

$$\hat{y}_t^{(p)} = L_{t-1} + hB_{t-1} + S_{t-1}, \text{ kde } h = 1, \dots, m \quad (2.72)$$

a intervalové předpovědi určíme obdobně jako u dvojitého exponenciálního vyrovnávání dle předpisu (2.55), kde d nahradíme d_h .

2.4.5. Testy náhodnosti

Pro posouzení vhodnosti modelu nám poslouží také testy náhodnosti, jež jsou schopné otestovat, zda rezidua, získaná na základě použitého modelu, vykazují pouze bílý šum. U všech testů se tedy zabýváme hypotézou H_0 , zda $e_t \sim WN(0, \sigma^2 I_n)$, oproti alternativě H_1 , že rezidua netvoří bílý šum. Díky těmto testům tak zjistíme, zda jsme nezapomněli zahrnout do modelu nějakou složku, jako např. špatně zachycený trend nebo opomenutí periodicky opakujících se změn. V rámci většího rozsahu výběru uvažujeme $n > 20$.

Poznámka 2.6. Tyto testy lze použít i na neočištěné ČŘ.

Test založený na znaménkách diferencí

V tomto testu zkoumáme počet kladných diferencí z řady reziduí e_1, \dots, e_n , který značíme k . Čili zajímá nás, ve kterých bodech řada roste. Jestliže jsou sousední hodnoty reziduí stejné, pak je až na jednu z nich vyškrtáme. Body, ve kterých řada roste, obdržíme pomocí nově definované náhodné veličiny V_t , která je dána předpisem:

$$V_t = \begin{cases} 1, & e_{t+1} > e_t \\ 0, & e_{t+1} < e_t \end{cases}.$$

Pro výpočet testové statistiky je nutné určit střední hodnotu a rozptyl počtu kladných diferencí k , přičemž předpokládáme, že pravděpodobnost růstu či poklesu hodnoty rezidua vzhledem k předchozí hodnotě odpovídá v obou případech $\frac{1}{2}$.

$$E(k) = E\left(\sum_{t=1}^{n-1} V_t\right) = \frac{n-1}{2}$$

a

$$\text{var}(k) = \text{var}\left(\sum_{t=1}^{n-1} V_t\right) = \frac{n+1}{12}. \quad [16]$$

Při větším rozsahu výběru zamítáme hypotézu H_0 přibližně na hladině spolehlivosti α , pokud

$$\frac{\left|k - \frac{n-1}{2}\right|}{\sqrt{\frac{n+1}{12}}} \geq u_{1-\frac{\alpha}{2}}.$$

Test založený na bodech zvratu

Jak název napovídá, naším úkolem bude nalézt takové body e_t , ve kterých se řada reziduí „láme“. Lokální minima, resp. maxima, odpovídají dolnímu, resp. hornímu bodu zvratu. Celkový počet bodů zvratu označme r . Pokud jsou sousední hodnoty reziduí stejné, pak je až na jednu z nich vyškrtáme. Pro výpočet střední

hodnoty a rozptylu počtů bodů zvratu definujeme náhodnou veličinu U_t , pro kterou platí:

$$U_t = \begin{cases} 1, & e_{t-1} < e_t > e_{t+1} \vee e_{t-1} > e_t < e_{t+1} \\ 0, & \text{jinak} \end{cases} .$$

Střední hodnotu a rozptyl náhodné veličiny r určíme následujícími vztahy

$$E(r) = E\left(\sum_{t=1}^{n-2} U_t\right) = \frac{2(n-2)}{3}$$

a

$$\text{var}(r) = \text{var}\left(\sum_{t=1}^{n-2} U_t\right) = \frac{16n-29}{90}. \quad [16] \quad (2.73)$$

Při větším rozsahu výběru zamítáme hypotézu H_0 přibližně na hladině spolehlivosti α , když

$$\frac{\left|r - \frac{2(n-2)}{3}\right|}{\sqrt{\frac{16n-29}{90}}} \geq u_{1-\frac{\alpha}{2}}.$$

Test založený na Kendallově koeficientu pořadové korelace τ

Pro tento test je nezbytné napočítat v takových dvojic y_s a y_t , pro které platí, že $y_s < y_t$ pro $s < t$. Dále spočítáme tzv. Kendallův koeficient pořadové korelace τ , který je dán vztahem

$$\tau = \frac{4v}{n(n-1)} - 1.$$

Tento koeficient mimo jiné vyjadřuje míru závislosti mezi různě uspořádanými hodnotami, přičemž $\tau \in \langle -1, 1 \rangle$. Jeho interpretace je obdobná korelačnímu koeficientu. Kendallův koeficient tak můžeme využít k testování nezávislosti reziduí prostřednictvím standardizované statistiky, ve které předpokládáme nulovou střední hodnotu a rozptyl $\text{var}(\tau) = \frac{2(2n+5)}{9n(n-1)}$.

Při větším rozsahu výběru zamítáme hypotézu H_0 přibližně na hladině významnosti α , pokud

$$\frac{|\tau|}{\sqrt{\frac{2(2n+5)}{9n(n-1)}}} \geq u_{1-\frac{\alpha}{2}}.$$

Test založený na Spearmanově koeficientu ρ

Nechť q_1, \dots, q_n značí pořadí hodnot dané řady (reziduí). Při shodě pořadí hodnot reziduí uvažujeme průměrné pořadí. Pak Spearmanův koeficient pořadové korelace ρ vypočítáme pomocí vztahu

$$\rho = 1 - \frac{6}{n(n^2 - 1)} \sum_{i=1}^n (i - q_i)^2.$$

Uvedený koeficient představuje neparametrickou verzi Pearsonova korelačního koeficientu. Pro větší rozsah výběru zamítáme hypotézu H_0 přibližně na hladině spolehlivosti α , jestliže

$$|\rho| \sqrt{n-1} \geq u_{1-\frac{\alpha}{2}}.$$

Mediánový test

Výpočet testovací statistiky je založen na výpočtu výběrového mediánu z ČŘ reziduí, jenž zapisujeme jako 0,5 kvantil $e_{0,5}$. Prakticky si to můžeme představit tak, že hledáme takový počet změn, kdy hodnota reziduí přesáhne nebo klesne pod hodnotu mediánu reziduí. Pokud by hodnota mediánu nerozdělila rezidua na dvě stejně početné skupiny (např. vlivem více stejných hodnot mediánu), přesuneme dané reziduum do méně početné skupiny. Ostatní rezidua, vykazující se shodnou hodnotou s mediánem, již nebudeme v testu uvažovat.

Nechť u představuje celkový počet změn přes hodnotu mediánu reziduí a m označuje celkový počet reziduí, jejichž hodnota je nižší, než hodnota mediánu reziduí. Pak při větším rozsahu reziduí zamítáme hypotézu H_0 přibližně na hladině spolehlivosti α , pokud

$$\frac{|u - (m + 1)|}{\sqrt{m(m - 1) / (2m - 1)}} \geq u_{1 - \frac{\alpha}{2}}.$$

Poznámka 2.7. Pro menší rozsahy porovnáváme hodnoty testovacích statistik s kritickými hodnotami ze speciálních tabulek, které lze najít např. v [13, 14, 16].

Pro zjištění, zda jsme neopomněli do modelu zahrnout lineární trend, nám poslouží test založený na znaménkách diferencí, Kendallově koeficientu τ a Spearmanově koeficientu ρ , oproti tomu opomenutí periodicity již tyto testy neodhalí. Při podezření na změny spíše periodické je vhodnější využít test založený na bodech zvratu nebo mediánový test.

3. Analýza klientských vkladů

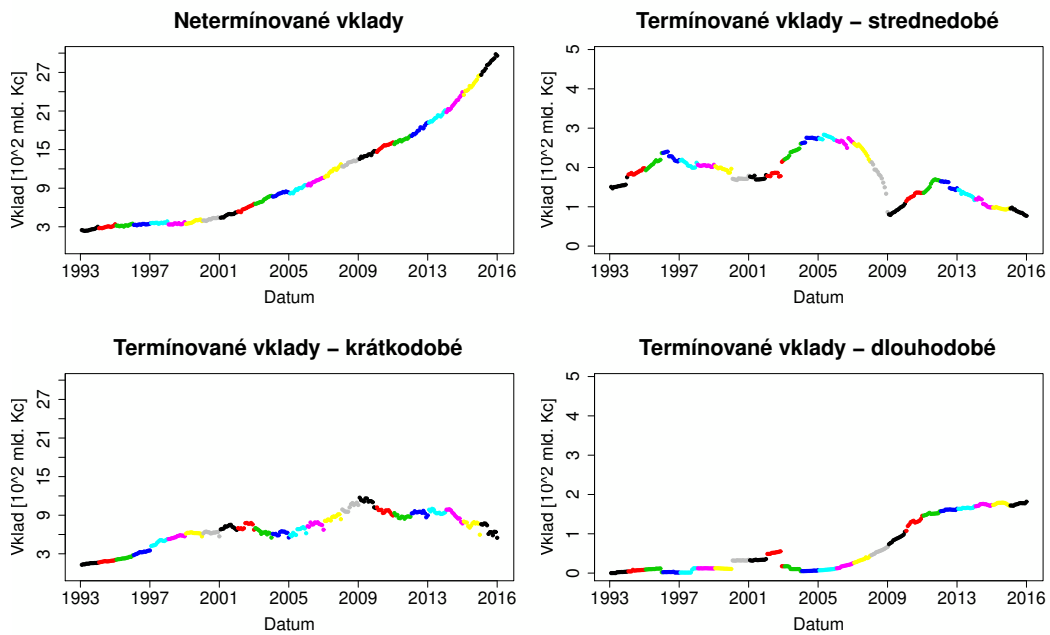
V této kapitole budeme analyzovat reálné datové soubory prostřednictvím dosud uvedených metod. Praktické výpočty provedeme ve statistickém softwaru R. Výhodou softwaru R je jeho jednoduchost a flexibilita. Díky nepřebornému množství balíčků můžeme na data aplikovat nejnovější statistické postupy. Pro analýzu využijeme měsíční data z České národní banky, jež se týkají uložených vkladů v komerčních bankách spravované ČNB v období 1993 - 2016. K dispozici máme na výběr netermínované vklady, krátkodobé, střednědobé a dlouhodobé termínované vklady. Naším cílem je kompletní analýza datových souborů s využitím různých modelů, které srovnáme z hlediska kvality modelu a přesnosti predikcí. Období 1993 - 2015 uvažujeme jako historické (referenční), rok 2016 jako predikční.

Veškeré zdrojové kódy, jež nebudou publikované přímo v textu, budou umístěné na konci diplomové práce nebo na příloženém CD.

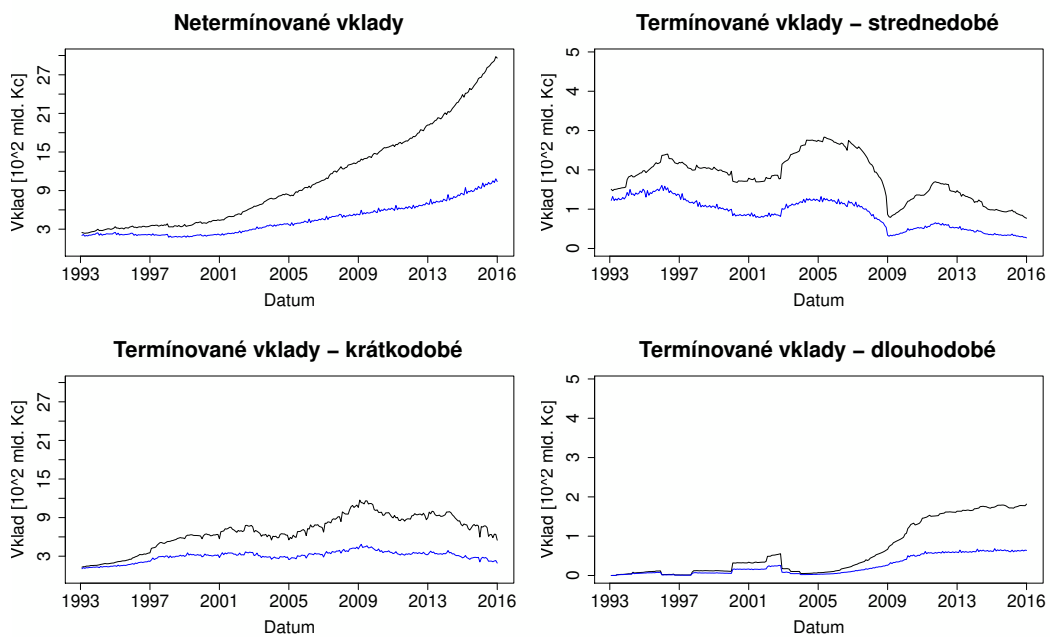
3.1. Úprava a načtení dat

Než se pustíme do samotné analýzy dat, nedílnou součástí analýzy představuje úprava a vizualizace dat. Naši pozornost bychom měli upřít také na zdroj a rozsah dat. Jelikož data pocházejí z databáze ČNB, můžeme je považovat za důvěryhodné. Výhodou těchto dat je, že neobsahují chybějící hodnoty a máme dostatečný rozsah souboru.

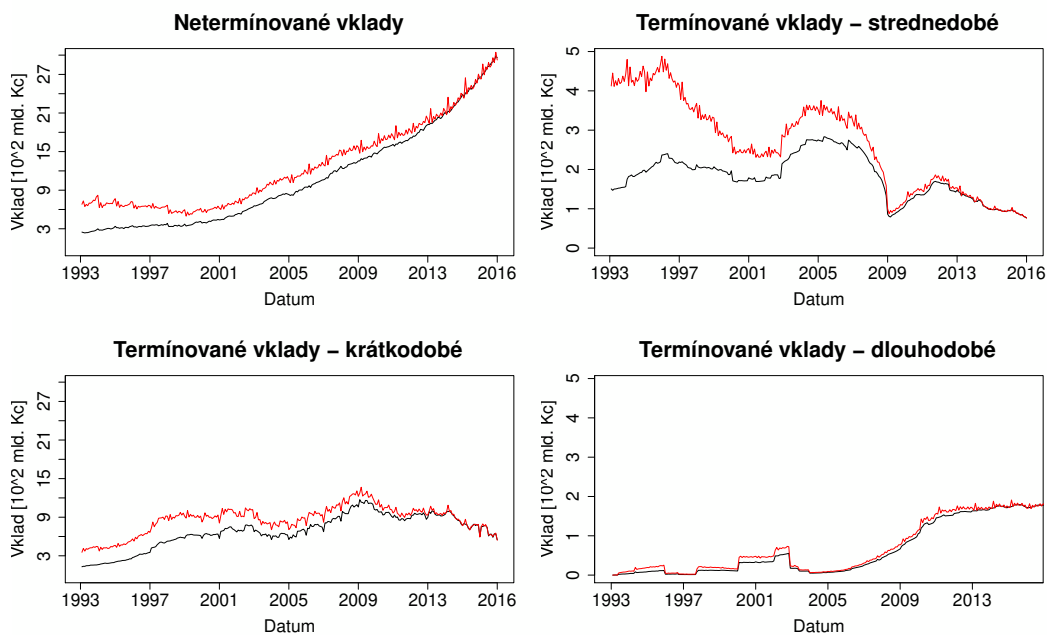
Když data vykreslíme (obrázek 2), můžeme na první pohled vidět vzrůstající tendenci netermínovaných vkladů. Naopak vklady krátkodobého charakteru po určitém růstu stagnovaly a následně docházelo k mírnému poklesu. Pro lepší orientaci jsou v obrázku 2 barevně odděleny hodnoty příslušící k jednotlivým rokům. Abychom lépe pochopili podstatu dat, měli bychom zvážit faktory, které mohou ovlivnit hodnoty výše vkladů. Mezi nejvýznamnější faktory tak můžeme zařadit míru inflace, hospodářské cykly, či jiné ekonomické a politické faktory. Na obrázcích 3, resp. 4 vidíme nezanedbatelný vliv inflace, kde původní hodnoty jsou zvýrazněny černou barvou a očištěné hodnoty barvou modrou, resp. červenou.



Obrázek 2: Data termínovaných a netermínovaných vkladů od České národní banky – roky jsou barevně odděleny.



Obrázek 3: Diskontování vkladů k roku 1993 – modrá křivka zobrazuje diskontované vklady k roku 1993, černá křivka znázorňuje původní data.



Obrázek 4: Úročení vkladů k roku 2016 – červená křivka zobrazuje úročené vklady k roku 2016, černá křivka znázorňuje původní data.

nou. Ať už ČŘ diskontujeme k datu 1993 nebo naopak úročíme k poslednímu roku, vidíme výrazné narovnání ČŘ. Tím dostaneme reálnější představu, zda v průběhu času skutečně došlo k navýšení, či snížení celkového objemu vkladů. Jako kosmetické, ale potřebné, se jeví očištění od kalendářních dnů, které vzhledem k vysokým hodnotám nemá na ČŘ tak výrazný vliv jako míra inflace. Pro naši další analýzu je vhodné pracovat s úročenými daty k poslednímu roku, neboť budeme predikovat hodnoty i k následujícímu roku při předpokládané inflaci následujícího roku. Kvůli nízkému zájmu o střednědobé a dlouhodobé termínované vklady, budeme dále porovnávat datové soubory netermínovaných a krátkodobých termínovaných vkladů očištěných od kalendářních dnů a úročených k roku 2016.

Data z ČNB jsou dostupné v tabulkovém programu Excel [21]. Pro načtení dat do softwaru R je potřeba data uložit do souboru s příponou .csv a načíst pomocí příkazu:

```
data = read.csv2("vklady_k2016.csv", header=TRUE)
```

Pro lepší představu datových souborů slouží tabulka 2, ve které je znázorněna

ukázka dat. V záhlaví dat si můžeme všimnout neočištěných a úročených dat k roku 2016 dle vzorce (2.3).

Po načtení dat si do proměnných uložíme následující údaje, které využíváme u všech metod:

```
datovy_soubor = "Neterminovane_vklady"
# případně "Kratkodobe_terminovane_vklady"
m=12 # typ dat (měsíční, čtvrtletní, roční, atd.)
alfa = 0.05

roky = length(unique(data[, "Rok"])) - 1
n = roky * m
N = length(data)
y=data[c(1:n), datovy_soubor]
Y=data[, datovy_soubor]
rok = as.factor(data[, "Rok"])
bud_cas=(n+1):(n+m)
bud_hodnoty=data[bud_cas, datovy_soubor]
t=1:n
```

Ověření normality jsme provedli prostřednictvím jednovýběrového Kolmogorova–Smirnova testu [17]. V tomto testu porovnáváme rezidua s normálním rozdělením se střední hodnotou 0 a rozptylem odpovídající výše rozptylu reziduí. Hypotéza H_0 odpovídá, zda se náhodný výběr neliší od vybraného teoretického rozdělení. Test spustíme příkazem:

```
ks.test(x, "pnorm", mean=0, sd = y) # Za x zvolíme rezidua,
# za y odpovídající směrodatnou odchylku spočítanou z reziduí.
```

U neterminovaných vkladů jsme zamítli hypotézu H_0 na hladině významnosti $\alpha = 0,05$ u lineárního, exponenciálního a logistického trendu. U ostatních modelů jsme nemohli zamítnout hypotézu H_0 , že rezidua pochází z normálního rozdělení.

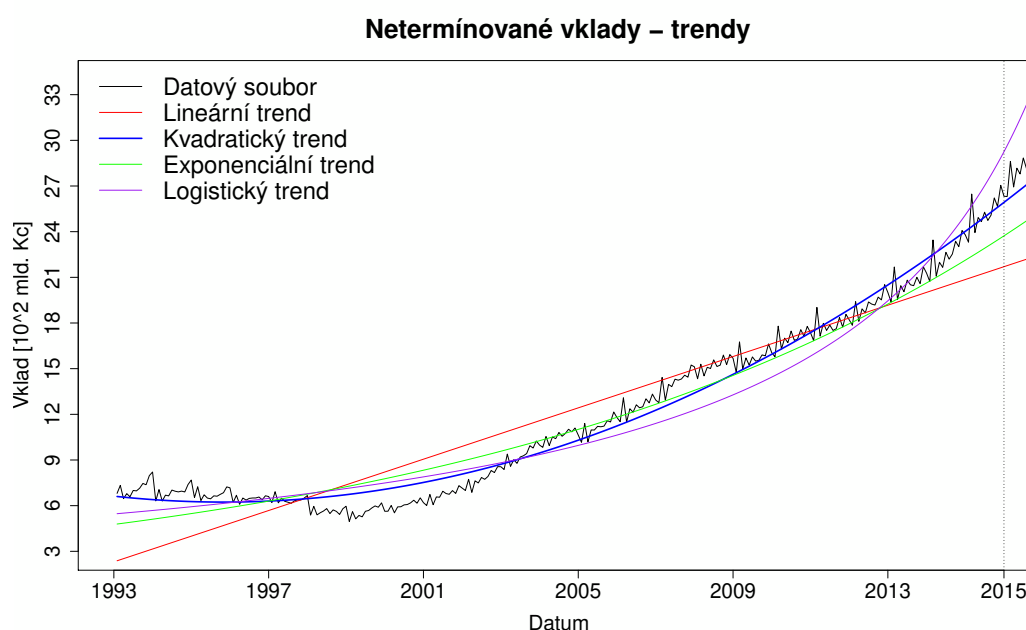
Tabulka 2: Ukázka datového souboru klientských vkladů [mil. Kč].

Datum	Rok	Netermínované vklady		Krátkodobé termínované vklady		Střednědobé termínované vklady		Dlouhodobé termínované vklady	
		původní	úročené	původní	úročené	původní	úročené	původní	úročené
31.01.1993	1993	248565	680820	129370	354345	150584	412450	21	58
28.02.1993	1993	241797	733224	136666	414436	146838	445282	85	257
31.03.1993	1993	235879	646074	136278	373264	150183	411351	154	427
30.04.1993	1993	239921	679048	145487	411772	150432	425767	31	87
31.05.1993	1993	240468	658641	146425	401058	151207	414156	2039	5584
30.06.1993	1993	247254	699802	147569	417664	152423	431402	2239	6338
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
31.10.2016	2016	2919282	2884400	642628	634950	81088	80120	177760	175636
30.11.2016	2016	2982198	3044782	639403	652821	78446	80093	177593	181320
31.12.2016	2016	2959795	2924428	548955	542396	76918	75999	181554	179385

V rámci krátkodobých termínovaných vkladů jsme zamítli H_0 pouze u exponenciálního trendu. Nyní již můžeme přejít k trendovým funkcím.

3.2. Porovnání trendových funkcí

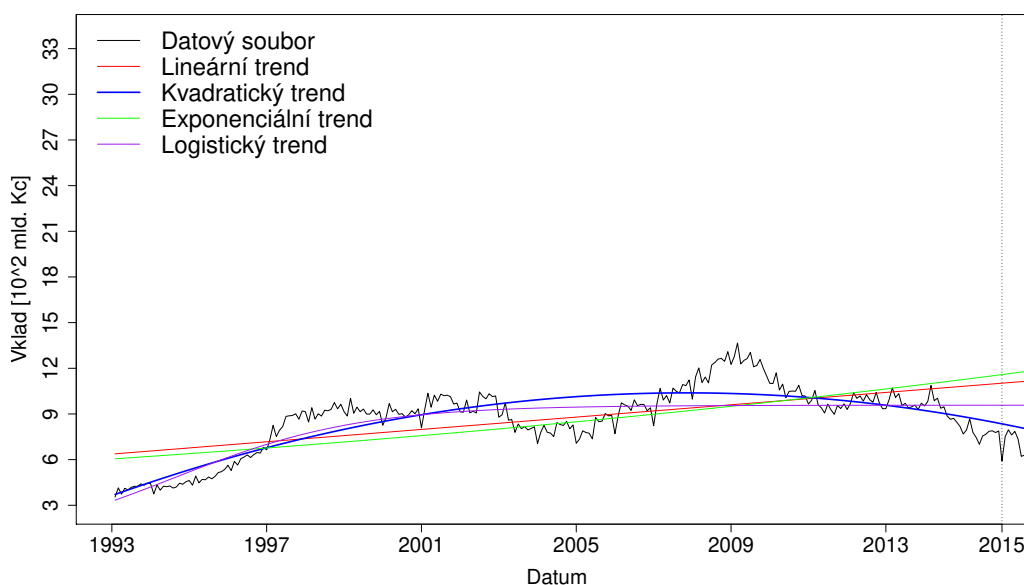
Trendy zkonstruujeme dle teorie z kapitoly 2.4.2. Již na základě vizualizace trendů z obr. 5 a 6 je patrná flexibilita kvadratického trendu, kterému se velmi daří popsat ČR.



Obrázek 5: Srovnání trendových funkcí u netermínovaných vkladů.

K objektivnějšímu srovnání trendových funkcí nám poslouží tabulky 3 a 4, ve kterých jsou zaznamenané charakteristiky jednotlivých trendových modelů. Pro posouzení vhodnosti modelu z hlediska historických dat nás bude zajímat index determinace. Z hlediska predikcí naopak MADP a MSEF. Neopomeneme ani komplexnější charakteristiky AIC a BIC, jež berou v úvahu i počet parametrů v modelu. Dle těchto charakteristik se nejlépe jeví kvadratický trend, který dominuje u obou typů charakteristik. Zároveň si však můžeme všimnout, že využití pouze trendových funkcí není vhodné pro všechny typy ČR. Dále můžeme předpokládat zlepšení těchto charakteristik, když k trendové složce připojíme se-

Krátkodobé termínované vklady – trendy



Obrázek 6: Srovnání trendových funkcí u termínovaných krátkodobých vkladů.

zónní nebo cyklickou složku (viz kapitola 3.3). Nejhorší výsledky jsme obdrželi pro logistický trend, jenž se zejména na data krátkodobě termínovaných vkladů vůbec nehodí.

Tabulka 3: Charakteristiky trendových funkcí u netermínovaných vkladů.

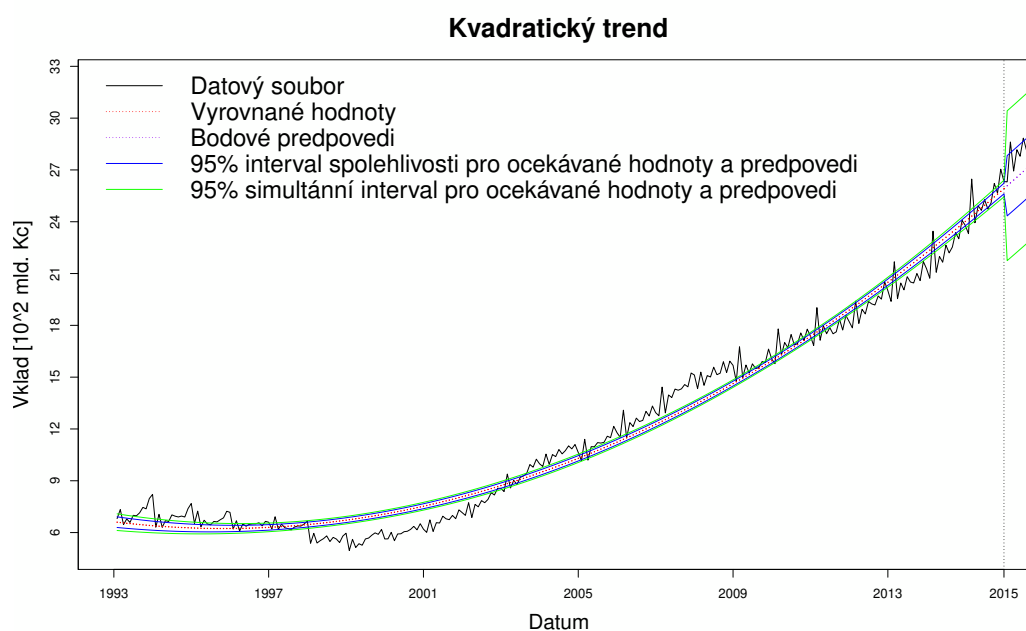
Trend Charakteristiky	R^2	AIC	BIC	MADP	RMSEP	Počet parametrů
Lineární	0,877	6770	6781	629342	635614	2
Kvadratický	0,979	6285	6299	145041	161308	3
Exponenciální	0,951	6515	6526	380552	387715	2
Logistický	0,942	6562	6573	374030	388448	3

Vzhledem k úspěšnosti kvadratického trendu a splnění normality, představíme intervalové odhady pouze pro kvadratický trend u obou datových souborů. Na obrázku 7, resp. 8 jsou znázorněny 95% oboustranné intervalové odhady pro očekávané hodnoty a předpovědi pro netermínované, resp. krátkodobé termínované vklady. Z těchto obrázků můžeme velmi dobře rozpoznat charakteristiku simul-

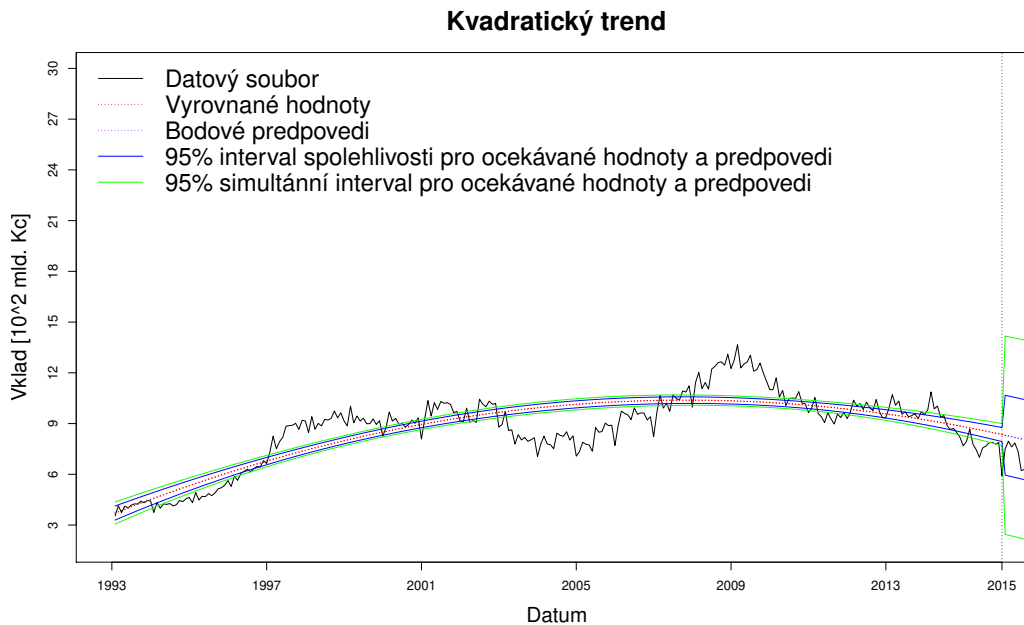
Tabulka 4: Charakteristiky trendových funkcí u krátkodobých termínovaných vkladů.

Trend Charakteristiky	R^2	AIC	BIC	MADP	RMSEP	Počet parametrů
Lineární	0,389	6649	6660	434853	443010	2
Kvadratický	0,703	6451	6466	128687	144958	3
Exponenciální	0,296	6688	6699	498481	506181	2
Logistický	0,67	6479	6490	279196	290304	3

tánních intervalů, kde v mediánu ČŘ pozorujeme nejužší rozpětí intervalu a naopak na koncích ČŘ se rozpětí intervalů rozšiřuje. Model vsutku dobře zachytil i budoucí vývoj ČŘ, neboť budoucí hodnoty jsou také zahrnuté v intervalových předpovědích.



Obrázek 7: 95% intervalové odhady kvadratického trendu u netermínovaných vkladů.



Obrázek 8: 95% intervalové odhady kvadratického trendu u krátkodobých termínovaných vkladů.

3.3. Periodické modely

Sezónní složku namodelujeme prostřednictvím modelů z teoretické části v kapitole 2.4.3. V tomto případě očistíme data současně od trendové i sezónní složky, a to s využitím modelu s umělými proměnnými a modelu skrytých period. Při očištění datových souborů od trendové složky využijeme poznatky z kapitoly 3.2. V obou případech datové soubory očistíme kvadratickým trendem, jež dosahoval výborných výsledků.

Nejdříve si popíšeme analýzu dat pomocí modelu s umělými proměnnými. Abychom dosáhli přiměřené složitosti modelu, rozhodli jsme na měsíční data obou datových souborů aplikovat čtvrtletní typ sezónnosti. Díky této skutečnosti odhadujeme v modelu celkem šest parametrů, přičemž k trendové funkci se vztahují tři parametry a další tři parametry odpovídají umělým proměnným. První čtvrtletí považujeme za referenční.

Model skrytých period, obsahující konstantní trend, jsme aplikovali na rezidua získaná očištěním datových souborů od kvadratického trendu, přičemž ve výsled-

ném modelu můžeme absolutní člen z MSP zanedbat. Při analýze významných frekvencí jsme našli mnoho významných frekvencí, kde ty nejvýznamnější frekvence byly dlouhodobějšího charakteru. Nicméně, neměla by nás odradit vysoká absence krátkodobých frekvencí, neboť i dlouhodobější frekvence mohou zkvalitnit navržený model. U datových souborů netermínovaných vkladů, resp. krátkodobých termínovaných vkladů, jsme zvolili první tři, resp. čtyři nejvýznamnější frekvence s délkou periody 11,5, 23 a 0,2 roku, resp. 11,5, 7,6, 5,75 a 3,8 roku. Konečný model včetně trendu má následující podobu:

$$y_t = \delta_0 + \delta_1 t + \delta t^2 + \sum_{j=1}^V \alpha_j \sin(\omega_j t) + \sum_{j=1}^V \beta_j \cos(\omega_j t) + \epsilon_t, \quad t = 1, \dots, n. \quad (3.1)$$

V rámci kvadratického trendu jsme obdrželi odhad parametrů δ u netermínovaných vkladů $\hat{\delta} = (662493, -2291, 34)^T$ se směrodatnou odchylkou jednotlivých složek vektoru $\hat{\delta}$ ve výši 15872, 267 a 0,92. Odhad parametrů δ u krátkodobých termínovaných vkladů je $\hat{\delta} = (363415,7577, -21)^T$ se směrodatnou odchylkou jednotlivých složek vektoru $\hat{\delta}$ v hodnotě 21466, 358 a 1,25.

V MSP, aplikovaného na rezidua, odhadujeme sedm, resp. devět parametrů. Odhad parametrů γ u netermínovaných vkladů je $\hat{\gamma} = (3,12 \cdot 10^{-10}, 83337,31206, -46326, -8670, -34204,1721)^T$ se směrodatnou odchylkou ve výši 2650 pro absolutní člen a 3748 pro ostatní složky vektoru parametrů $\hat{\gamma}$. Odhad parametrů γ u krátkodobých termínovaných vkladů je $\hat{\gamma} = (3,19 \cdot 10^{-9}, -12113, -116125, -3296,57699, -53202, -5327,46244, -18096)^T$ se směrodatnou odchylkou ve výši 3149 pro absolutní člen a 4453 pro ostatní složky vektoru parametrů $\hat{\gamma}$. Jak jsme již dříve zmínili, jelikož je výše absolutního členu v MSP v obou případech blízká nule, můžeme jej zanedbat.

Poznámka 3.1. K analýze počtu významných frekvencí jsme použili i Sieglouvu modifikaci [5, 15]. Díky této metodě jsme zjistili šest, resp. čtyři významné frekvence u netermínovaných, resp. krátkodobých termínovaných vkladů. Jelikož tato informace nevedla ke zlepšení již navrhnutého modelu, nebudeme se Sieglouvu modifikací zabývat.

V tabulkách 5, resp. 6 jsou charakteristiky modelu s umělými proměnnými a modelu skrytých period pro datové soubory netermínovaných, resp. krátkodobě termínovaných vkladů.

Tabulka 5: Charakteristiky sezónních modelů u netermínovaných vkladů.

Model Charakteristiky	R^2	AIC	BIC	MADP	RMSEP	Počet parametrů
Model s umělými proměnnými	0,979	6287	6313	145756	160860	6
Model skrytých period	0,995	5915	5955	106911	119577	9

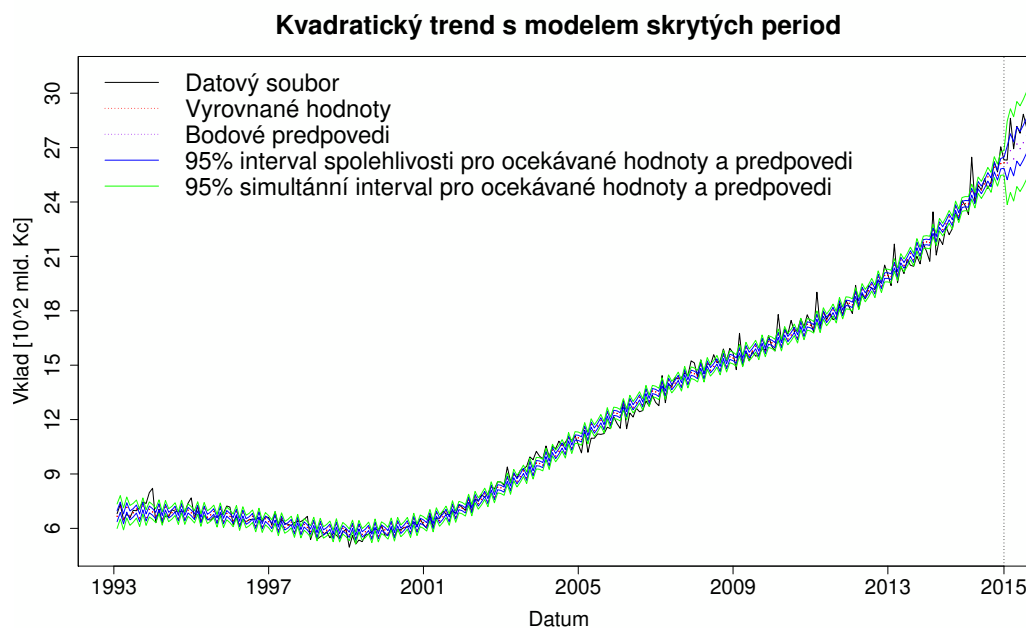
Tabulka 6: Charakteristiky sezónních modelů u krátkodobých termínovaných vkladů.

Model Charakteristiky	R^2	AIC	BIC	MADP	RMSEP	Počet parametrů
Model s umělými proměnnými	0,706	6455	6480	129339	144047	6
Model skrytých period	0,943	6010	6053	72871	88294	11

Můžeme postřehnout, že u obou datových souborů má model s kvadratickým trendem a umělými proměnnými téměř stejné charakteristiky jako kvadratický trend, tudíž model nezachytil žádné výkyvy krátkodobého charakteru. Naopak vzrostla složitost modelu. Z tohoto důvodu jsou hodnoty AIC a BIC u kvadratického trendu o trochu nižší, než u modelu s kvadratickým trendem a umělými proměnnými, což svědčí ve prospěch modelu s trendovou funkcí. Oproti tomu model skrytých period s kvadratickým trendem je dle charakteristik výrazně lepší, jelikož dokáže zachytit i výkyvy dlouhodobějšího charakteru. Model skrytých period vyniká nejenom v kvalitě modelu, ale také v přesnosti predikcí. Jedinou nevýhodou tohoto modelu je jeho složitost, neboť, jak jsme již zmínili, pracujeme dohromady s devíti, resp. s jedenácti parametry.

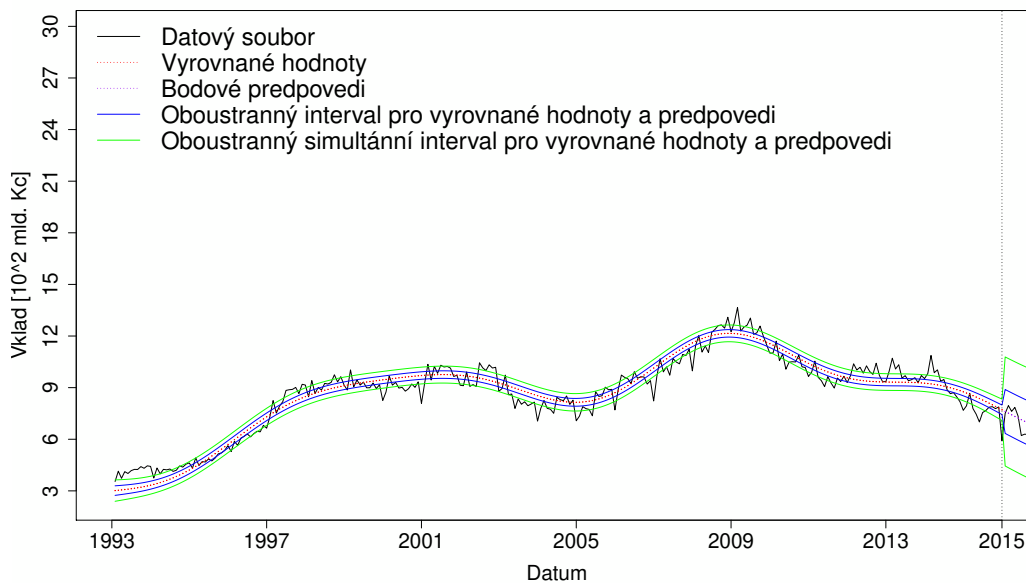
Jelikož jsme nezamítli H_0 na hladině $\alpha = 0,05$, že rezidua z modelu (3.1) pochází z normálního rozdělení (Kolmogorův-Smirnovův test), můžeme sestro-

jit 95% oboustranné intervalové odhady modelu skrytých period s kvadratickým trendem pro netermínované, resp. krátkodobé termínované vklady, které jsou znázorněny na obrázcích 9, resp. 10. Všimněme si, že oboustranné předpovědní intervaly v tomto modelu opět obsahují budoucí hodnoty datových souborů. Dále se nabízí srovnání intervalových odhadů sezónních a trendových modelů. Z obrázků 7 a 9 je patrné, že předpovědní oboustranné intervaly jsou u modelu skrytých period s kvadratickým trendem výrazně užší, než když periodickou složku neuvažujeme. To samé můžeme pozorovat i u krátkodobých termínovaných vkladů (obrázky 8 a 10). Ve prospěch modelu skrytých period s kvadratickým trendem svědčí nejen charakteristiky kvality modelu (R^2 , AIC, BIC), ale také přesnost bodových předpovědí (charakteristiky MADP a RMSEP).



Obrázek 9: 95% intervalové odhady modelu skrytých period s kvadratickým trendem u netermínovaných vkladů.

Kvadratický trend s modelem skrytých period



Obrázek 10: 95% intervalové odhady modelu skrytých period s kvadratickým trendem u krátkodobých termínovaných vkladů.

3.4. Adaptivní metody

Alternativu k očištění trendové a periodické složky představují adaptivní metody, které jsme představili v kapitole 2.4.4. V rámci diplomové práce uvedeme očištění datových souborů prostřednictvím dvojitého exponenciálního vyrovnávání a aditivní Holtovy-Wintersovy metody.

Dvojitě exponenciální vyrovnávání probíhá ve dvou fázích, kdy v první fázi zjistíme optimální hodnotu vyhlazovací konstanty α . V části druhé spočítáme pomocí vyrovnávacích statistik vyrovnané hodnoty a bodové předpovědi. V našem případě jsme v první fázi určili počáteční hodnoty vyrovnávacích statistik s využitím MNČ odhadu parametrů $\hat{\beta}_0^0$ a $\hat{\beta}_1^0$, které jsme spočítali pro prvních šest záznamů ČR. Dále jsme postupovali dle teorie z kapitoly 2.4.4. Výstupem první fáze je optimální vyhlazovací konstanta v hodnotě $\alpha = 0,87$, resp. $\alpha = 0,75$ pro netermínované, resp. krátkodobé termínované vklady. V druhé části jsme zvolili počáteční hodnoty vyrovnávacích statistik obdobnou formou, tedy prostřednictvím MNČ odhadu lineárního trendu, ovšem z první poloviny ČR. Tím

jsme dosáhli přesnějšího odhadu počátečních hodnot startovacích statistik, jelikož máme velmi obsáhlou ČR.

V rámci Holtovy-Wintersovy metody jsme zvolili optimální hodnoty vyrovnávacích konstant na základě článku [12] ve výši $\alpha = 0,3$ pro úroveň trendu, $\beta = 0,1$ vztahující se ke směrnici trendu a $\gamma = 0,3$ pro sezónní složku. Počáteční hodnoty úrovně trendu L_0 , směrnice trendu B_0 a vektoru sezónní složky S_0 jsme zvolili na základě MNČ odhadu z modelu s umělými proměnnými včetně lineárního trendu pro čtvrtletní sezónu. Obdrželi jsme počáteční hodnoty ve výši $L_0 = 7021$, $B_0 = 231526$ a $S_0 = (0,0,0,5943,5943,5943,8385,8385,8385,26777,26777,26777)^T$, kde složky vektoru S_0 odpovídají jednotlivým měsícům v roce, přičemž první čtvrtletí je referenční. V tabulkách 7 a 8 představíme charakteristiky obou adaptivních modelů. Vysoká kvalita modelu, jež popisuje historická data, je doménou těchto přizpůsobujících se modelů v čase. Proto nás bude zajímat především kvalita predikce adaptivních metod u obou datových souborů. Bodové předpovědi jsou dle charakteristik MADP a RMSEP podstatně lepší než v případě trendových funkcí. Velmi zajímavé je srovnání kvality predikcí adaptivních a sezónních metod. Obě již zmíněné adaptivní metody si jeví oproti předchozím modelům v kapitole 3.3 velmi dobře. Nicméně musíme vyzdvihnout Holtovu-Wintersovu metodu, která vyniká v přesnosti predikcí oproti ostatním modelům. Možné zlepšení Holtovy-Wintersovy metody závisí i na subjektivní volbě vyrovnávacích parametrů.

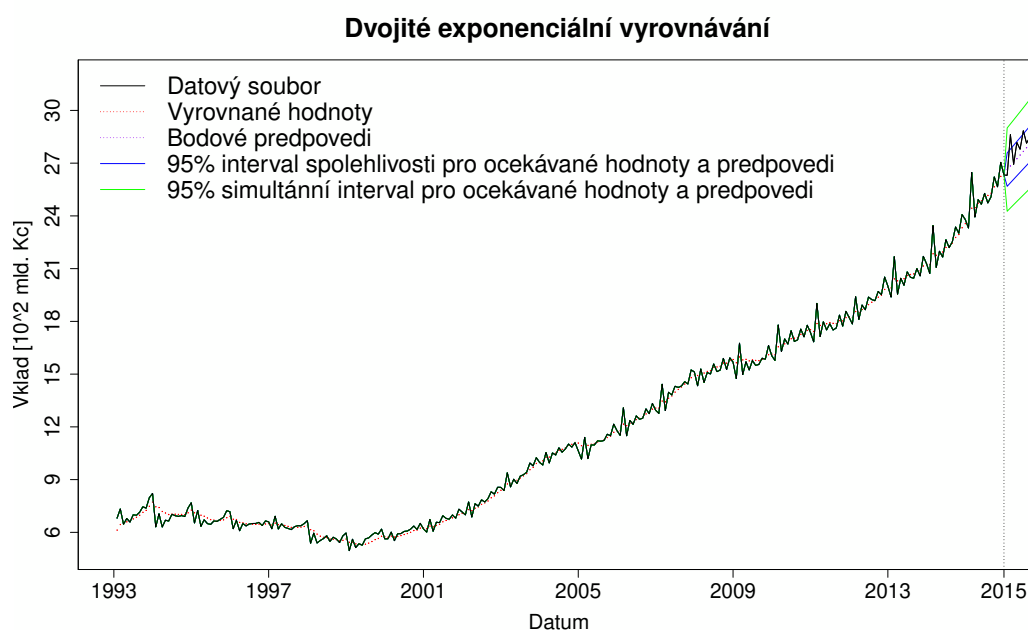
Tabulka 7: Charakteristiky adaptivních modelů u netermínovaných vkladů.

Model Charakteristiky	R^2	RSC	RMSE	MADP	RMSEP
Dvojité exponenciální vyrovnávání	0,996	$41,42 \cdot 10^{10}$	38741	74969	94787
Holtova-Wintersova metoda	0,998	$15,99 \cdot 10^{10}$	24068	67311	72575

Nezbývá nám, než si představit 95% oboustranné předpovědní intervaly dvojitěho exponenciálního vyrovnávání a Holtovy-Wintersovy metody u obou da-

Tabulka 8: Charakteristiky adaptivních modelů u krátkodobých termínovaných vkladů.

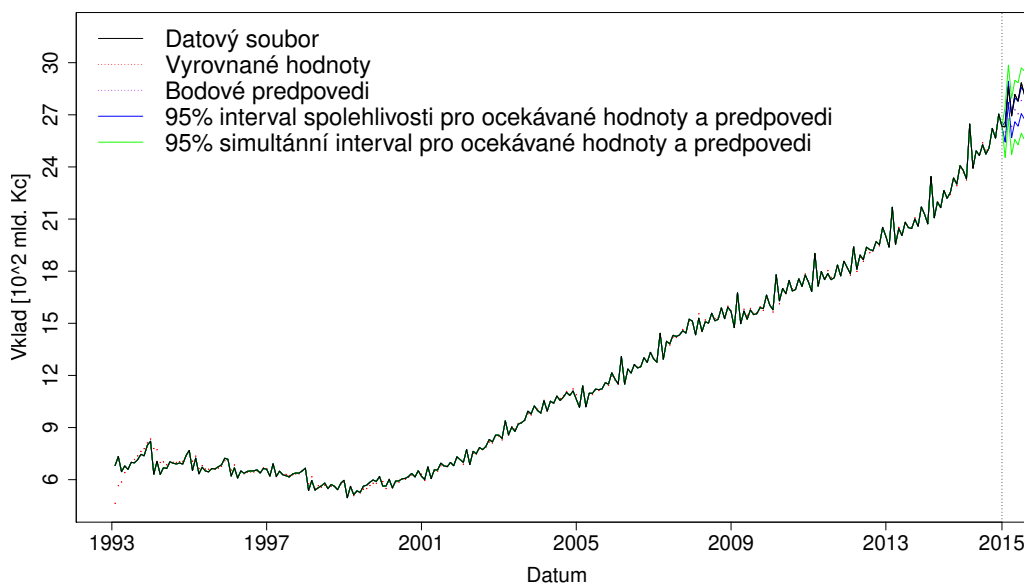
Model	Charakteristiky	R^2	RSČ	RMSE	MADP	RMSEP
Dvojitě exponenciální vyrovnávání		0,983	$22,24 \cdot 10^{10}$	28386	73436	86375
Holtova-Wintersova metoda		0,988	$14,77 \cdot 10^{10}$	23134	41898	54203



Obrázek 11: 95% předpovědní intervaly dvojitěho exponenciálního vyrovnávání u netermínovaných vkladů.

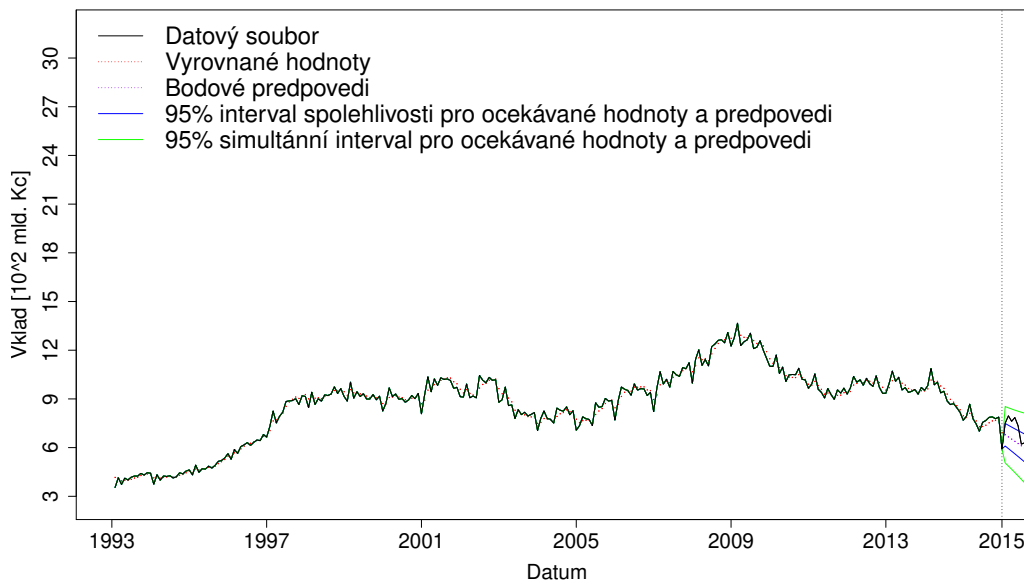
ových souborů znázorněné na obrázcích 11 - 14. Předností těchto předpovědních intervalů je jejich přesnost, neboť opět obsahují skutečné budoucí hodnoty datových souborů. Užší předpovědní intervaly můžeme vidět u Holtovy-Wintersovy metody.

Holtova–Wintersova metoda

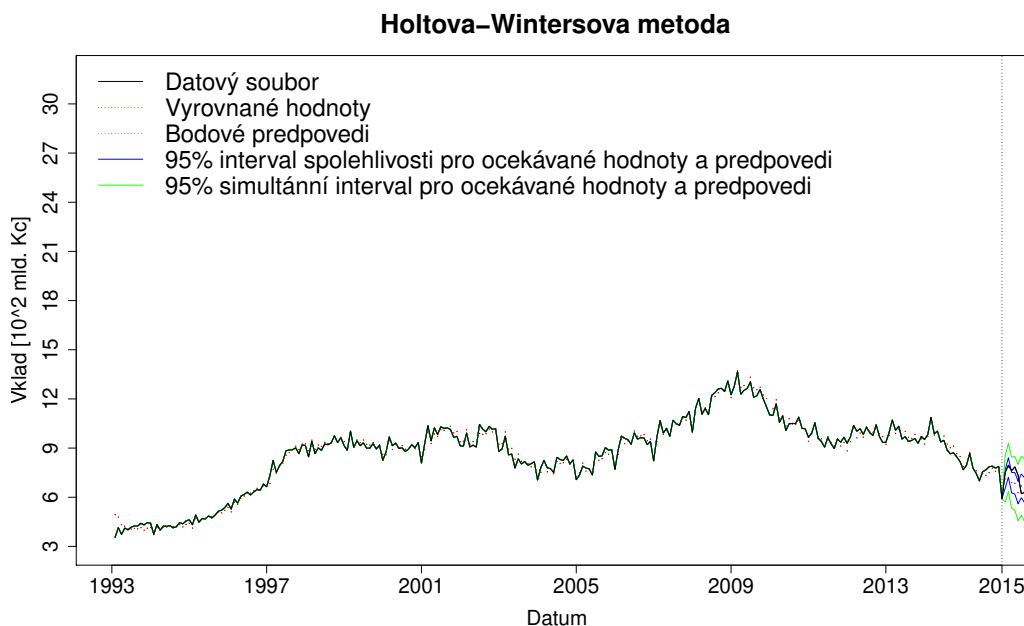


Obrázek 12: 95% předpovědní intervaly Holtovy-Wintersovy metody u netermínovaných vkladů.

Dvojitě exponenciální vyrovnávání



Obrázek 13: 95% předpovědní intervaly dvojitě exponenciálního vyrovnávání u krátkodobých termínovaných vkladů.

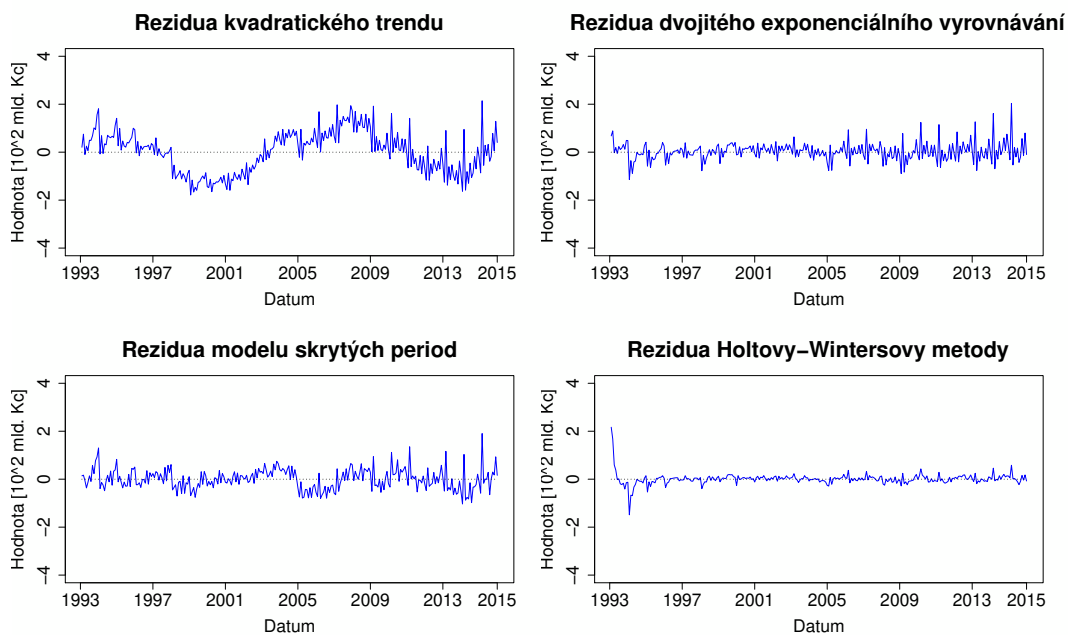


Obrázek 14: 95% předpovědní intervaly Holtovy-Wintersovy metody u krátkodobých termínovaných vkladů.

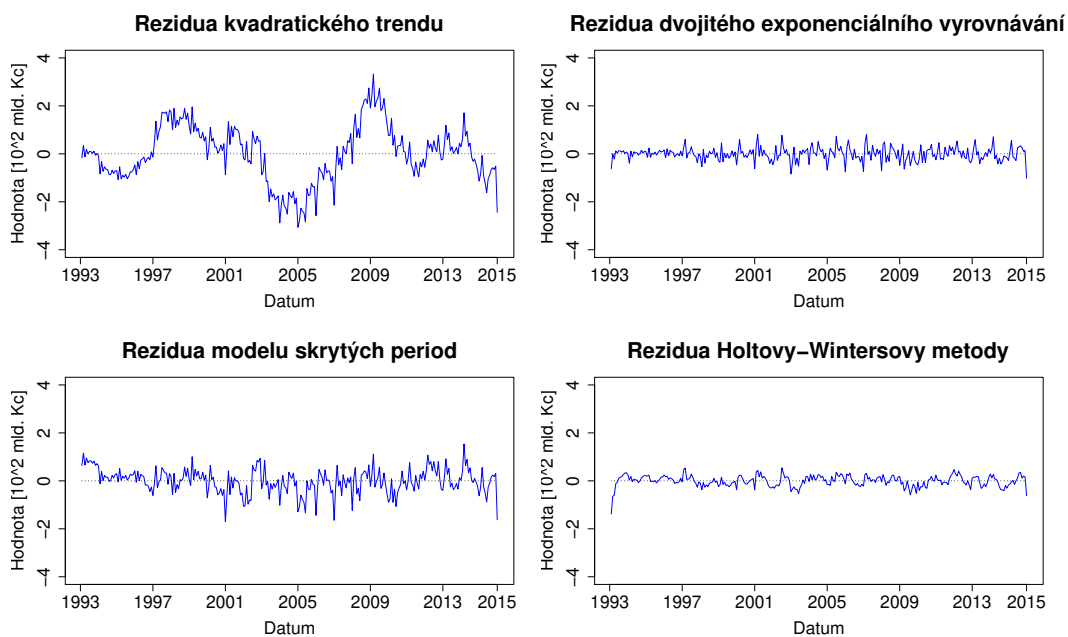
3.5. Identifikace náhodné složky

Ke zjištění, zda jsme v modelech neopomněli žádnou složku, využijeme testy náhodnosti z kapitoly 2.4.5, které si představíme na reziduích kvadratického trendu, u modelu skrytých period s kvadratickým trendem, u dvojitého exponenciálního vyrovnavání a u Holtovy-Wintersovy metody. Vykreslená rezidua jsou znázorněná na obrázcích 15 a 16.

Když se podíváme na rezidua kvadratického trendu, můžeme si všimnout zjevné absence periodické složky u obou datových souborů. Jak jsme již analyzovali v kapitole 3.3, jedná se o pravidelné výkyvy dlouhodobějšího charakteru. Můžeme se domnívat, že ony výkyvy jsou způsobeny hospodářským cyklem, kdy se střídají fáze konjunktury a recese. Velmi dobře je patrný výrazný nárůst ukládaných vkladů od trendu v letech 2008 - 2009, kdy vypukla ekonomická krize způsobená hypoteční krizí v USA. Můžeme se tak domnívat, že ekonomické subjekty byly velmi opatrné a reagovaly na tuto krizi výraznějším ukládáním svých finančních prostředků, naopak v následujících letech, kdy již nastává konjunktura, mů-



Obrázek 15: Rezidua u netermínovaných vkladů.



Obrázek 16: Rezidua u krátkodobých termínovaných vkladů.

žeme předpokládat jisté uvolnění ekonomických subjektů. Tehdy dochází ke snižování výše uložených vkladů od trendu, a tudíž i k vyšší spotřebě těchto zdrojů.

Nyní přejdeme k analýze náhodné složky. Pro lepší orientaci si sestavíme tabulky testů, do kterých umístíme informace o zamítnutí, či nezamítnutí hypotézy H_0 . Zamítnutí H_0 svědčí ve prospěch opomenutí některé složky. V případě, že nelze zamítnout hypotézu H_0 , budeme se domnívat, že model velmi dobře popsal datový soubor a zbylá rezidua představují bílý šum. Výsledky testů pro netermínované, resp. krátkodobé termínované vklady jsou znázorněné v tabulce 9, resp. v tabulce 10.

Z tabulek 9 a 10 je patrné, že přizpůsobivost adaptivních metod vede k reziduům, které dle většiny testů náhodnosti můžeme označit jako bílý šum. Špatně si nevede ani model skrytých period s kvadratickým trendem, u kterého jsme nemohli zamítnout hypotézu H_0 ve třech, resp. ve dvou případech v rámci netermínovaných, resp. krátkodobých termínovaných vkladů. Také si můžeme všimnout velké citlivosti mediánového testu, jež zamítal H_0 téměř ve všech případech.

Tabulka 9: Testy náhodnosti – rozhodnutí o hypotéze H_0 u netermínovaných vkladů na hladině $\alpha = 0,05$.

Test Model	Model skrytých period	Dvojité exponenciální vyrovnávání	Holtova-Wintersova metoda
Test založený na znaménkách diferencí	nelze zamítnout	nelze zamítnout	nelze zamítnout
Test založený na bodech zvratu	zamítáme	zamítáme	nelze zamítnout
Test založený na Kendallově koeficientu pořadové korelace τ	nelze zamítnout	nelze zamítnout	nelze zamítnout
Test založený na Spearmanově koeficientu	nelze zamítnout	nelze zamítnout	nelze zamítnout
Mediánový test	zamítáme	zamítáme	zamítáme

Tabulka 10: Testy náhodnosti – rozhodnutí o hypotéze H_0 u krátkodobých termínovaných vkladů na hladině $\alpha = 0,05$.

Test Model	Model skrytých period	Dvojitě exponenciální vyrovnávání	Holtova-Wintersova metoda
Test založený na znaménkách diferencí	nelze zamítnout	nelze zamítnout	nelze zamítnout
Test založený na bodech zvratu	nelze zamítnout	zamítáme	zamítáme
Test založený na Kendallově koeficientu pořadové korelace τ	zamítáme	nelze zamítnout	nelze zamítnout
Test založený na Spearmanově koeficientu	zamítáme	nelze zamítnout	nelze zamítnout
Mediánový test	zamítáme	nelze zamítnout	zamítáme

Závěr

Tato práce se zabývala analýzou klientských vkladů z let 1993 až 2016, a to pomocí nástrojů dekompozičního přístupu k časovým řadám. Jedná se o přístup, který se využívá u datových souborů, jehož hodnoty souvisí s časovým obdobím.

Práci můžeme rozdělit na tři hlavní části. V první části práce jsem nastínil ekonomické souvislosti týkající se klientských vkladů. V části druhé jsem uvedl teoretické aspekty týkající se statistické analýzy časových řad. Zejména jsem se pak zaměřil na dekompoziční přístup, který rozkládá časovou řadu na čtyři složky - trendovou, sezónní, cyklickou a náhodnou. V rámci první složky jsem se zabýval trendovými funkcemi a dvojitým exponenciálním vyrovnáním, jež charakterizují dlouhodobější změny časových řad. Sezónní a cyklickou složku jsem popisoval prostřednictvím modelu skrytých period. Také jsem zmínil model s umělými proměnnými, který zachycuje pouze sezónní změny během jednotlivých let. V rámci adaptivních metod zmiňuji i Holtovu-Wintersovu metodu, která je zobecněním exponenciálního vyrovnání, ve smyslu zahrnutí jak trendové, tak i sezónní složky. V rámci čtvrté složky jsem zmínil testy náhodnosti, které signalizují opomenutí některé ze složek v modelu.

V třetí části práce jsem analyzoval již zmíněné klientské vklady prostřednictvím teorie z předchozí kapitoly pomocí statistického softwaru R. V rámci analýzy jsem porovnal jednotlivé metody, a to jak z hlediska charakteristik týkajících se kvality modelu, tak dle charakteristik souvisejících s přesností predikovaných hodnot výše uložených klientských vkladů. Mezi nejúspěšnější modely můžeme zařadit model skrytých period s kvadratickým trendem, jenž odhalil dlouhodobé odchylky od trendu. Dlouhodobější odchylky tak kopírují vývoj hospodářského cyklu, během kterého se mění poptávka klientů k uložení svých vkladů. Zejména můžeme zaznamenat výraznější ukládání klientských vkladů během hospodářské krize a značnější spotřebu těchto vkladů během hospodářské konjunktury. Dále bych rád vyzdvihnul Holtovu-Wintersovu metodu, jelikož dosahovala vysoké přesnosti bodových i intervalových předpovědí. V rámci trendové složky vynikal model založený na kvadratické funkci a dvojitě exponenciální vyrovnávání. Zmíněné

modely jsou doplněny o grafy s bodovými a intervalovými odhady.

Při psaní této práce jsem nabyl zkušenosti z oblasti statistické analýzy datových souborů, a to zejména díky odborné literatuře a zdokonalení dovedností v softwaru R. Věřím, že tato práce bude názornou ukázkou, jak obdobné datové soubory můžeme analyzovat.

Literatura

- [1] Česká národní banka. Metodický list - Vklady klientské. [cit. 15. 3. 2017 20:27]. Dostupné z: www.cnb.cz/docs/ARADY/MET_LIST/tvklob_cs.pdf.
- [2] PTÁČEK, J. Úspory a vklady. 2. aktualizované vyd. Praha: Bankovní institut, 1999. iii, 208 s. ISBN 80-7265-047-5.
- [3] Business.center.cz. Slovník pojmů. [cit. 20. 3. 2017 15:26]. Dostupné z: <http://business.center.cz/business/pojmy/>.
- [4] Zákon č. 586/1992 Sb., o daních z příjmů. §4. ÚZ 770. Vyd. Ostrava: Sagit, 2010. 272 s. ISBN 978-80-7208-786-0.
- [5] Cipra, T. Analýza časových řad s aplikacemi v ekonomii: Vysokoškol. učeb. pro stud. matem.-fyz. fak. 1. vyd. Praha, 1986.
- [6] Anděl, J. Statistická analýza časových řad. 1. vyd. Praha, 1976.
- [7] Montgomery, D. Forecasting and Time Series Analysis. 2nd Ed. New York: McGraw-Hill, 1990. 11, 381 s. ISBN 0-07-042858-1.
- [8] Hebák, P. a kol. Statistické myšlení a nástroje analýzy dat. 2. vydání. Praha: Informatorium, 2015. 877 stran. ISBN 978-80-7333-118-4.
- [9] Hyndman, R. J., Koehler, A. B., Ord J. K., Snyder, R. D. Prediction Intervals for Exponential Smoothing Using Two New Classes of State Space Models. [cit. 27. 11. 2017 11:27]. Dostupné z: <http://onlinelibrary.wiley.com/doi/10.1002/for.938/pdf>.
- [10] Tratar, L. F., Strmčnik, E. The comparison of Holt–Winters method and Multiple regression method: A case study. Journal Energy, 2016. ISSN 03605442. Dostupné z: https://ac.elscdn.com/S0360544216305357/1-s2.0S0360544216305357main.pdf?_tid=d9ad2d06b73111e7bc03-00000aab0f6b&acdnat=1508681219_e2c71001688cdb9616616cc698d475a1.

- [11] Bermúdez, J. D., Segura, J.V., Vercher E. A decision support system methodology for forecasting of time series based on soft computing. *Comput Stat Data Anal*, 2006. str. 5117791.
- [12] Chatfield, Ch., Yar, M.. Holt-Winters Forecasting: Some Practical Issues. *Journal of the Royal Statistical Society. Series D (The Statistician)*, vol. 37, no. 2, 1988, pp. 129140. JSTOR, dostupné: www.jstor.org/stable/2348687.
- [13] Veterinární a farmaceutická univerzita Brno, Statistické tabulky. [cit. 25. 11. 2017 20:25]. Dostupné z: <https://cit.vfu.cz/statpotr/POTR/Teorie/tabulky.htm>.
- [14] University of Washington. TABLES. [cit. 25. 11. 2017 20:27]. Dostupné z: <http://faculty.washington.edu/heagerty/Books/Biostatistics/TABLES/>.
- [15] Siegel, A. F. Testing for Periodicity in a Time Series. *Journal of the American Statistical Association*, vol. 75, č. 370, 1980, str. 345–348. [cit. 8. 12. 2017 20:27] Dostupné z: www.jstor.org/stable/2287456.
- [16] Moore, G. H. , Wallis, W. A. Time Series Significance Tests Based on Signs of Differences. *Journal of the American Statistical Association*, Vol. 38, č. 222. 1943. [cit. 8. 12. 2017 21:27] Dostupné z: <https://www.jstor.org/stable/pdf/2279535.pdf>
- [17] Taylor A., Kolmogorov-Smirnov Tests. [cit. 9. 12. 2017 20:47] Dostupné z: <https://www.rdocumentation.org/packages/dgof/versions/1.2/topics/ks.test>

Zdroje dat

- [18] Česká spořitelna. Profil České spořitelny. [cit. 23. 11. 2017 15:47]. Dostupné z: www.csas.cz/cs/o-nas/kdo-jsme.

- [19] ČSOB, Presentace výsledků. [cit. 23. 11. 2017 15:50]. Dostupné z:
www.csob.cz/portal/documents/10710/460146/csob-presentace-vysledku-4q2016-cz.pdf.
- [20] Komerční banka. Tisková zpráva. [cit. 23. 11. 2017 15:53]. Dostupné z:
www.kb.cz/cs/o-bance/tiskove-centrum/tiskove-zpravy/kb-v-roce-2016-prispevek-k-rozvoji-ceske-a-slovenske-ekonomiky-zvysovani-efektivnosti-a-pripravy-na-budoucnost-1275/.
- [21] Česká národní banka. Databáze časových řad ARAD. [cit. 25. 11. 2017 21:15].
Dostupné z: https://www.cnb.cz/cnb/STAT.ARADY_PKG.PARAMETRY_SESTAVY?p_sestuid13278&p_strid=AABCAB&p_langCS.

A. Trendová složka

A.1. Lineární trend

```
X_lt = matrix(data=c(rep(1,n),1:n),ncol=2,nrow = n)
beta_lt = solve(t(X_lt)%*%X_lt)%*%t(X_lt)%*%y
vyrovnane_lt = X_lt %*% beta_lt
e_lt=y-vyrovnane_lt

p_lt=length(beta_lt)
k_lt = p_lt-1
RSC_lt = sum(e_lt^2)
sigma2_lt = RSC_lt/(n-p_lt)
sigma_lt = sqrt(sigma2_lt)
St_lt = sum((y-mean(y))^2)
R_lt = 1-(RSC_lt/St_lt)
Ra_lt = 1-(1-R_lt)* (n+k_lt-p_lt)/(n-p_lt)

# oboustranný interval spolehlivosti

f_lt1=c()
polsirka_lt1 = c()
dolnihr_lt1 = c()
hornihr_lt1 = c()
sirka_lt1 = c()

for(i in 1:n){
  f_lt1[i] = t(X_lt[i,])%*(solve(t(X_lt)%*%X_lt))%*%X_lt[i,]
  polsirka_lt1[i] = sqrt(sigma2_lt * f_lt1[i])
  *qt(0.975,n-p_lt)
  dolnihr_lt1[i]=vyrovnane_lt[i]-polsirka_lt1[i]
```

```

hornihr_lt1[i]=vyrovnane_lt[i]+polsirka_lt1[i]
sirka_lt1[i]=2*polsirka_lt1[i]
}

# oboustranný simultánní interval spolehlivosti (pás spolehlivosti)

f_sim_lt1=c()
polsirka_sim_lt1=c()
dolnihr_sim_lt1=c()
hornihr_sim_lt1=c()
sirka_sim_lt1=c()

for( i in 1:n){
  f_sim_lt1[i] = t(X_lt[i,])%*%solve(t(X_lt)%*%X_lt)
  %*% X_lt[i,]
  polsirka_sim_lt1[i]= sigma_lt*sqrt(p_lt
  *qf(.975, df1=p_lt, df2=(n-p_lt))*f_sim_lt1[i] )
  dolnihr_sim_lt1[i] = vyrovnane_lt[i] - polsirka_sim_lt1[i]
  hornihr_sim_lt1[i] = vyrovnane_lt[i] + polsirka_sim_lt1[i]
  sirka_sim_lt1[i] = 2*polsirka_sim_lt1[i]
}

# bodové a intervalové předpovědi

x_p_lt = matrix(data=c(rep(1,m),bud_cas),ncol=2,nrow = m)
bod_predp_lt = x_p_lt %*%beta_lt

SEP = sum((bud_hodnoty-bod_predp_lt)^2)
MSEP_lt = sum((bud_hodnoty-bod_predp_lt)^2)/m
RMSEP_lt = sqrt(MSEP_lt)

```

```

MADP_lt = sum(abs(bud_hodnoty-bod_predp_lt))/m

k=2
AIC_lt = n*log(RSC_lt/n)+k*(p_lt+1)
BIC_lt = n*log(RSC_lt/n)+log(n)*(p_lt+1)

# předpovědní oboustranný interval

f_lt2=c()
polsirka_lt2 = c()
dolnihr_lt2 = c()
hornihr_lt2 = c()
sirka_lt2 = c()

for(i in 1:m){
  f_lt2[i] = 1 + t(x_p_lt[i,])%*(solve(t(X_lt)
    %*%X_lt))%*%x_p_lt[i,]
  polsirka_lt2[i] = sqrt(sigma2_lt * f_lt2[i])
    *qt(0.975,n-p_lt)
  dolnihr_lt2[i]=bod_predp_lt[i]-polsirka_lt2[i]
  hornihr_lt2[i]=bod_predp_lt[i]+polsirka_lt2[i]
  sirka_lt2[i]=2*polsirka_lt2[i]
}

# předpovědní oboustranný simultánní interval

f_sim_lt2=c()
polsirka_sim_lt2=c()
dolnihr_sim_lt2=c()
hornihr_sim_lt2=c()

```

```

sirka_sim_lt2=c()

for( i in 1:m){
  f_sim_lt2[i] = 1 + t(x_p_lt[i,])%%solve(t(X_lt)%%X_lt)
  %% x_p_lt[i,]
  polsirka_sim_lt2[i] = sigma_lt*sqrt(m
    *qf(.975, df1=m, df2=(n-p_lt)) * f_sim_lt2[i] )
  dolnihr_sim_lt2[i] = bod_predp_lt[i] - polsirka_sim_lt2[i]
  hornihr_sim_lt2[i] = bod_predp_lt[i] + polsirka_sim_lt2[i]
  sirka_sim_lt2[i] = 2*polsirka_sim_lt2[i]
}

```

A.2. Kvadratický trend

```

X_kt = matrix(data=c(rep(1,n),1:n, (1:n)^2),ncol=3,nrow = n)
beta_kt = solve(t(X_kt)%%X_kt)%%t(X_kt)%%y
vyrovnane_kt = X_kt %% beta_kt
e_kt=y-vyrovnane_kt

```

Ostatní charakteristiky, bodové a intervalové odhady obdržíme obdobně jako v příloze A.1, jen příponu "lt" nahradíme příponou "kt", přičemž

```

x_p_kt = matrix(data=c(rep(1,m),bud_cas,bud_cas^2),ncol=3,
  nrow = m)

```

A.3. Exponenciální trend

```

ln_y = log(y, base = exp(1))
X_et = matrix(data=c(rep(1,n),1:n),ncol=2,nrow = n)
beta_et = solve(t(X_et)%%X_et)%%t(X_et)%%ln_y
alfa_et = exp(beta_et[1])
gama_et = exp(beta_et[2])
vyrovnane_et = alfa_et * (gama_et^t)

```

Běžné charakteristiky obdržíme obdobně jako v příloze A.1. Příponu "lt" nahradíme příponou "et". Blíže si popíšeme změny ve výpočtu intervalových odhadů, které odhadneme na základě následujících charakteristik, zbytek kódu je obdobný dle A.1.

```
RSC_et_log= sum((ln_y-X_et**beta_et)^2)
sigma2_et_log = RSC_et_log/(n-p_et)
sigma_et_log=sqrt(sigma2_et_log)
```

```
# bodové předpovědi
```

```
x_p_et = matrix(data=c(rep(1,m),bud_cas),ncol=2,nrow = m)
bod_predp_et = alfa_et * (gama_et^bud_cas)
# bod_predp_et2 = exp(x_p_et**beta_et) # alternativa
```

Změny v rámci intervalových odhadů:

```
polsirka_et1[i] = sqrt(sigma2_et_log
  * f_et1[i]) * qt(0.975,n-p_et)
dolnihr_et1[i]= exp(X_et[i,] **beta_et
  - polsirka_et1[i])
hornihr_et1[i]= exp(X_et[i,] **beta_et
  + polsirka_et1[i])

polsirka_sim_et1[i]= sigma_et_log*sqrt(p_et
  *qf(.975, df1=p_et, df2=(n-p_et)) *f_sim_et1[i] )
dolnihr_sim_et1[i] = exp(X_et[i,] **beta_et
  - polsirka_sim_et1[i])
hornihr_sim_et1[i] = exp(X_et[i,] **beta_et
  + polsirka_sim_et1[i])
```

```

polsirka_et2[i] = qt(0.975,n-p_et)*sigma_et_log
                *sqrt(f_et2[i])
dolnihr_et2[i]=exp(x_p_et[i,] **beta_et
- polsirka_et2[i])
hornihr_et2[i]=exp(x_p_et[i,] **beta_et
+ polsirka_et2[i])

polsirka_sim_et2[i] = sigma_et_log*sqrt(m
    *qf(.975, df1=m, df2=(n-p_et)) * f_sim_et2[i] )
dolnihr_sim_et2[i] = exp(x_p_et[i,] **beta_et
- polsirka_sim_et2[i])
hornihr_sim_et2[i] = exp(x_p_et[i,] **beta_et
+ polsirka_sim_et2[i])

```

A.4. Logistický trend

```

g=ifelse(datovy_soubor=="Neterminovane_vklady",0.999,0.966)
X_logt = matrix(data=c(rep(1,n),g^t),ncol=2,nrow = n)
beta_logt = solve(t(X_logt)**X_logt)**t(X_logt)**(1/y)
delta2 = beta_logt[1]
alfa2 = beta_logt[2]
delta = 1/delta2
alfa = alfa2*delta

vyrovnane_logt = delta/ (1 + alfa * (g^t) )
# vyrovnane_logt2 = 1/(X_logt**beta_logt) # alternativa

```

Ostatní charakteristiky a bodové odhady obdržíme obdobně jako v příloze A.1. Příponu "lt" nahradíme příponou "logt", přičemž bodové předpovědi obdržíme

na základě následujícího kódu. Intervalové odhady neuvádíme, jelikož převrácené hodnoty nevykazují známky normality.

```
x_p_logt = matrix(data=c(rep(1,m),g^bud_cas),ncol=2,nrow = m)
bod_predp_logt = delta/ (1 + alfa * (g^bud_cas) )
# bod_predp_logt2=1/(x_p_logt%*%beta_logt) # alternativa
```

B. Periodická složka

B.1. Model s umělými proměnnými

```
p = roky # roky
r = 4 # sezony
z2 = c(0,0,0,1,1,1,0,0,0,0,0,0)
z3 = c(0,0,0,0,0,0,1,1,1,0,0,0)
z4 = c(0,0,0,0,0,0,0,0,0,1,1,1)

X_MsUP = matrix(c(rep(1,n),1:n,(1:n)^2,rep(z2,p),rep(z3,p),
rep(z4,p)),ncol=6,nrow=n)
beta_MsUP = solve(t(X_MsUP)%*%X_MsUP)%*%t(X_MsUP)%*%y
p_MsUP=length(beta_MsUP)

beta0 = beta_MsUP[1]
beta1 = beta_MsUP[2]
beta2 = beta_MsUP[3]
alfa1 = 0
alfa2 = beta_MsUP[4]
alfa3 = beta_MsUP[5]
alfa4 = beta_MsUP[6]
alfa_prum = mean(c(alfa1,beta_MsUP[4:6]))
```

```

S1 = alfa1 - alfa_prum
S2 = alfa2 - alfa_prum
S3 = alfa3 - alfa_prum
S4 = alfa4 - alfa_prum
beta00 = beta0 + alfa_prum

z1 = rep(c(1,1,1,0,0,0,0,0,0,0,0,0,0),p)
t = 1:n
vyrovnane_MsUP = beta00 + beta1 *t+ beta2 *(t^2) + S2
                * z2 + S3 * z3 + S4 * z4 - alfa_prum * z1

```

Ostatní charakteristiky a intervalové odhady obdržíme obdobně jako v příloze A.1. Příponu "lt" nahradíme příponou "MsUP", přičemž bodové odhady obdržíme následujícím kódem.

```

z11 = c(1,1,1,0,0,0,0,0,0,0,0,0)
x_p_MsUP=matrix(c(rep(1,m), (n+1):(n+m), ((n+1):(n+m))^2, z2, z3, z4),
                ncol=6, nrow=m)
bod_predp_MsUP = beta00 + beta1 *bud_cas+ beta2
                *(bud_cas^2) + S2 * z2 + S3 * z3 + S4 * z4 - alfa_prum * z11

```

B.2. Model skrytých period s kvadratickým trendem

```

# Nejprve namodelujeme rezidua kvadratického trendu
omega = c() # frekvence
for (j in 1:(n/2)){omega[j] = (2*pi*j/n)}
H=n/2
I = c() # periodogram
for (i in 1:H){I[i] = (1/(4*pi))*((2/n) * (sum(e_kt
                *sin(omega[i]*t)))^2 + (2/n)
                * (sum(e_kt* cos(omega[i]*t)))^2)}

```

```

I_sort = sort(I,decreasing = TRUE) # I seradime sestupne
YY = c() #vybereme prvni nejvetsi frekvenci a normujeme
for(j in 1:H){ YY[j]=I_sort[j]/sum(I_sort[j:H])}

H = n/2 # 138 # Testujeme, zda se jedná o významné frekvence
p_value_1 = H * (1 - YY[1])^(H-1)
H = n/2-1 # 137
p_value_2 = H * (1 - YY[2])^(H-1) # atd.

vyznamne_frekvence=order(I,decreasing = TRUE) # uspořádáme frekvence

# Matice plánů pro modelování reziduí kvadratického trendu
X_e_kt_MSP1 = matrix(c(rep(1,n),sin(omega[vyznamne_frekvence[1]]
    *(1:n)),cos(omega[vyznamne_frekvence[1]]*(1:n)),
sin(omega[vyznamne_frekvence[2]]*(1:n)),
cos(omega[vyznamne_frekvence[2]]*(1:n)),
sin(omega[vyznamne_frekvence[3]]*(1:n)),
cos(omega[vyznamne_frekvence[3]]*(1:n))),ncol=7,nrow=n)

X_e_kt_MSP2 = matrix(c(rep(1,n),sin(omega[vyznamne_frekvence[1]]
    *(1:n)),cos(omega[vyznamne_frekvence[1]]*(1:n)),
sin(omega[vyznamne_frekvence[2]]*(1:n)),
cos(omega[vyznamne_frekvence[2]]*(1:n)),
sin(omega[vyznamne_frekvence[3]]*(1:n)),
cos(omega[vyznamne_frekvence[3]]*(1:n)),
sin(omega[vyznamne_frekvence[4]]*(1:n)),
cos(omega[vyznamne_frekvence[4]]*(1:n))),ncol=9,nrow=n)

if (datovy_soubor=="Neterminovane_vklady") {
  X_e_kt_MSP=X_e_kt_MSP1} else {X_e_kt_MSP=X_e_kt_MSP2}

```

```

beta_e_kt_MSP=solve(t(X_e_kt_MSP)%*%X_e_kt_MSP)
                %*%t(X_e_kt_MSP)%*%e_kt

# Vyrovnané hodnoty MSP s kvadratickým trendem
vyrovnane_kt_MSP= vyrovnane_kt + X_e_kt_MSP%*%beta_e_kt_MSP

# Matice plánů pro modelování intervalů spolehlivosti

X_kt_MSP1=matrix(c(rep(1,n),(1:n),(1:n)^2,
                sin(omega[vyznamne_frekvence[1]]*(1:n)),
                cos(omega[vyznamne_frekvence[1]]*(1:n)),
                sin(omega[vyznamne_frekvence[2]]*(1:n)),
                cos(omega[vyznamne_frekvence[2]]*(1:n)),
                sin(omega[vyznamne_frekvence[3]]*(1:n)),
                cos(omega[vyznamne_frekvence[3]]*(1:n)) ),ncol=9,nrow=n)

X_kt_MSP2=matrix(c(rep(1,n),(1:n),(1:n)^2,
                sin(omega[vyznamne_frekvence[1]]*(1:n)),
                cos(omega[vyznamne_frekvence[1]]*(1:n)),
                sin(omega[vyznamne_frekvence[2]]*(1:n)),
                cos(omega[vyznamne_frekvence[2]]*(1:n)),
                sin(omega[vyznamne_frekvence[3]]*(1:n)),
                cos(omega[vyznamne_frekvence[3]]*(1:n)),
                sin(omega[vyznamne_frekvence[4]]*(1:n)),
                cos(omega[vyznamne_frekvence[4]]*(1:n)) ),ncol=11,nrow=n)

if (datovy_soubor=="Neterminovane_vklady") {
    X_kt_MSP=X_kt_MSP1} else {X_kt_MSP=X_kt_MSP2}

```

```

# Předpovědní matice plánů pro modelování reziduí kvadr. trendu
x_p_e_kt_MSP1 = matrix(c(rep(1,m),
  sin(omega[vyznamne_frekvence[1]]*bud_cas),
  cos(omega[vyznamne_frekvence[1]]*bud_cas),
  sin(omega[vyznamne_frekvence[2]]*bud_cas),
  cos(omega[vyznamne_frekvence[2]]*bud_cas),
  sin(omega[vyznamne_frekvence[3]]*bud_cas),
  cos(omega[vyznamne_frekvence[3]]*bud_cas)),ncol=7,nrow=m)

x_p_e_kt_MSP2 = matrix(c(rep(1,m),
  sin(omega[vyznamne_frekvence[1]]*bud_cas),
  cos(omega[vyznamne_frekvence[1]]*bud_cas),
  sin(omega[vyznamne_frekvence[2]]*bud_cas),
  cos(omega[vyznamne_frekvence[2]]*bud_cas),
  sin(omega[vyznamne_frekvence[3]]*bud_cas),
  cos(omega[vyznamne_frekvence[3]]*bud_cas),
  sin(omega[vyznamne_frekvence[4]]*bud_cas),
  cos(omega[vyznamne_frekvence[4]]*bud_cas) ),ncol=9,nrow=m)

if (datovy_soubor=="Neterminovane_vklady") {
  x_p_e_kt_MSP=x_p_e_kt_MSP1} else {x_p_e_kt_MSP=x_p_e_kt_MSP2}

# Matice plánů pro modelování intervalových předpovědí

x_p_kt_MSP1 = matrix(c(rep(1,m),bud_cas,bud_cas^2,
  sin(omega[vyznamne_frekvence[1]]*bud_cas),
  cos(omega[vyznamne_frekvence[1]]*bud_cas),
  sin(omega[vyznamne_frekvence[2]]*bud_cas),
  cos(omega[vyznamne_frekvence[2]]*bud_cas),
  sin(omega[vyznamne_frekvence[3]]*bud_cas),

```

```

cos(omega[vyznamne_frekvence[3]]*bud_cas) ),ncol=9,nrow=m)

x_p_kt_MSP2 = matrix(c(rep(1,m),bud_cas,bud_cas^2,
  sin(omega[vyznamne_frekvence[1]]*bud_cas),
  cos(omega[vyznamne_frekvence[1]]*bud_cas),
  sin(omega[vyznamne_frekvence[2]]*bud_cas),
  cos(omega[vyznamne_frekvence[2]]*bud_cas),
  sin(omega[vyznamne_frekvence[3]]*bud_cas),
  cos(omega[vyznamne_frekvence[3]]*bud_cas),
  sin(omega[vyznamne_frekvence[4]]*bud_cas),
  cos(omega[vyznamne_frekvence[4]]*bud_cas) ),ncol=11,nrow=m)

if (datovy_soubor=="Neterminovane_vklady") {
  x_p_kt_MSP=x_p_kt_MSP1} else {x_p_kt_MSP=x_p_kt_MSP2}

# Bodové předpovědi MSP s kvadratickým trendem
bod_predp_kt_MSP= bod_predp_kt + x_p_e_kt_MSP %*%beta_e_kt_MSP

Ostatní charakteristiky a intervalové odhady obdržíme obdobně jako v příloze
A.1. Příponu "lt" nahradíme příponou "kt_MSP".

```

C. Adaptivní metody

C.1. Dvojitě exponenciální vyrovnávání

```

# Druhá fáze

alfa_exv2=ifelse(datovy_soubor=="Neterminovane_vklady",
  0.87,0.75) # již optimální alfa

```

```

beta_poc_exv2 = lm(y[1:(n/2)]~c(1:(n/2)))
S1_poc_exv2 = beta_poc_exv2$coefficients[1]
               -beta_poc_exv2$coefficients[2]*alfa_exv2/(1-alfa_exv2)
S2_poc_exv2 = beta_poc_exv2$coefficients[1]
               -beta_poc_exv2$coefficients[2]*2*alfa_exv2/(1-alfa_exv2)

S1_exv2=c()
S2_exv2=c()
for (i in 2:n ){
  S1_exv2[1]= (1-alfa_exv2)*y[1] + alfa_exv2*S1_poc_exv2
  S2_exv2[1]= (1-alfa_exv2)*S1_exv2[1] + alfa_exv2*S2_poc_exv2
  S1_exv2[i]= (1-alfa_exv2)*y[i] + alfa_exv2*S1_exv2[i-1]
  S2_exv2[i]= (1-alfa_exv2)*S1_exv2[i] + alfa_exv2*S2_exv2[i-1]
}

beta1_exv2=2*S1_exv2-S2_exv2
beta2_exv2= (1-alfa_exv2)/alfa_exv2 *(S1_exv2-S2_exv2)

vyrovnane_exv2 = beta1_exv2 + beta2_exv2*0

```

Ostatní charakteristiky a předpovědní intervalové odhady obdržíme obdobně jako v příloze A.1. Příponu "lt" nahradíme příponou "exv2" a místo směrodatné odchylky počítáme s RMSE, bodové předpovědi vypočítáme dle následujícího kódu.

```

bod_predp_exv2=c()
for (i in 1:m ){
  bod_predp_exv2[i] = beta1_exv2[n] + beta2_exv2[n]*i
}

```

C.2. Holtova-Wintersova metoda

alfa=0.3

beta=0.1

gama=0.3

a0_poc_W = beta00_1

b1_poc_W = beta1_1

S_poc_1_W = 0

S_poc_2_W = alfa2_1

S_poc_3_W = alfa3_1

S_poc_4_W = alfa4_1

sez_poc_W = c(rep(S_poc_1_W,3),rep(S_poc_2_W,3),
rep(S_poc_3_W,3),rep(S_poc_4_W,3))

a0_W=c()

b1_W=c()

sez_W=c()

for (i in 2:m){

a0_W[1] = alfa*(y[1] - S_poc_1_W) +(1-alfa)
*(a0_poc_W +b1_poc_W)

b1_W[1]= beta *(a0_W[1] - a0_poc_W) + (1-beta)
*b1_poc_W

sez_W[1] = gama * (y[1]-a0_W[1]) + (1-gama)
*sez_poc_W[1]

a0_W[i] = alfa*(y[i] - sez_poc_W[i]) +(1-alfa)
*(a0_W[i-1]+b1_W[i-1])


```

b1_W[i]= beta *(a0_W[i] - a0_W[i-1]) + (1-beta)
    *b1_W[i-1]
sez_W[i] = gama * (y[i]-a0_W[i]) + (1-gama) * sez_poc_W[i]

for (j in (1+m):n){
    a0_W[j] = alfa*(y[j] - sez_W[j-m]) +(1-alfa)
    *(a0_W[j-1]+b1_W[j-1])
    b1_W[j]= beta *(a0_W[j] - a0_W[j-1]) + (1-beta)
    *b1_W[j-1]
    sez_W[j] = gama * (y[j]-a0_W[j]) + (1-gama)
    *sez_W[j-m]
}}

vyrovnane_W=a0_W+sez_W

```

Ostatní charakteristiky a předpovědní intervalové odhady obdržíme obdobně jako v příloze C.1 a bodové předpovědi vypočítáme dle následujícího kódu.

```

bod_predp_W=c()
for (i in 1:m){
    bod_predp_W[i] = a0_W[n]+b1_W[n]*i+sez_W[n-m+i]
}

```

D. Testy náhodnosti

```
pocet=8 # počet typů reziduí
rezidua = matrix(c(e_lt,e_kt,e_et,e_logt,e_MsUP,
  e_kt_MSP,e_exv2,e_W), ncol=pocet, nrow=n )
```

D.1. Test založený na znaménkách diferencí

```
k_zd=matrix(0,nrow=n-1, ncol=pocet)
kladne_zd=c()
zaporne_zd=c()
test_zd=c()
rozhodnuti_zd=c()

for (i in 1:pocet){
  for (j in 1:(n-1)){
    k_zd[j,i] = ifelse((rezidua[j+1,i] > rezidua[j,i]),1,0) }
kladne_zd[i]=sum(k_zd[,i])
zaporne_zd[i]=(n-1)-sum(k_zd[,i])
test_zd[i]=abs((kladne_zd[i]-((n-1)/2)))/sqrt((n+1)/12)
rozhodnuti_zd[i]=ifelse(test_zd[i]>qnorm(0.975),
  "H0 zamítáme", "H0 nelze zamítnout")}
```

D.2. Test založený na bodech zvratu

```
r_bz=matrix(0,nrow=n-2, ncol=pocet)
kladne_bz=c()
zaporne_bz=c()
test_bz=c()
rozhodnuti_bz=c()
```

```

for (i in 1:pocet){
  for (j in 1:(n-2)){
    r_bz[j,i] = ifelse((rezidua[j+1,i] > rezidua[j,i])
    & (rezidua[j+2,i] < rezidua[j+1,i]) , 1,
    ifelse((rezidua[j+1,i] < rezidua[j,i])
& (rezidua[j+2,i] > rezidua[j+1,i]) , 1, 0 ))}
    kladne_bz[i]=sum(r_bz[,i])
    zaporne_bz[i]=(n-2)-sum(r_bz[,i])
    test_bz[i]=abs(kladne_bz[i]-(2*(n-2)/3))
    /sqrt((16*n-29)/90)
    rozhodnuti_bz[i]=ifelse(test_bz[i]>qnorm(0.975),
    "H0 zamítáme", "H0 nelze zamítnout")
  }
}

```

D.3. Test založený na Kendallově koeficientu pořadové korelace τ

```

v_ken=rep(0,pocet)
tau_ken=c()
test_ken=c()
rozhodnuti_ken=c()

for (i in 1:pocet){
  for (j in 1:(n-1)){
    v_ken[i] = v_ken[i] + sum(rezidua[j,i] < rezidua[(j+1):n,i])}
    tau_ken[i]=((v_ken[i]*4)/(n*(n-1)))-1
    test_ken[i]=abs(tau_ken[i])/ (sqrt((2*(2*n+5))
    /(9*n*(n-1))))
    rozhodnuti_ken[i]=ifelse(test_ken[i]>qnorm(0.975),
    "H0 zamítáme", "H0 nelze zamítnout")
  }
}

```

D.4. Test založený na Spearmanově koeficientu ρ

```
rho_sk=c()
test_sk=c()
rozhodnuti_sk=c()

for (i in 1:pocet){
  rho_sk[i] = 1 - (6/(n*(n^2 - 1)))*sum((1:n
    -rank(rezidua[,i], ties.method="average"))^2)
  test_sk[i] = abs(rho_sk[i])*sqrt(n-1)
  rozhodnuti_sk[i]=ifelse(test_sk[i]>qnorm(0.975),
    "H0 zamítáme", "H0 nelze zamítnout")}
```

D.5. Mediánový test

```
rozdil_med=matrix(0,ncol=pocet, nrow=n)
u_med=c()
m_med=c()
test_med=c()
rozhodnuti_med=c()

for (i in 1:pocet){
  rozdil_med[,i]=rezidua[,i]-median(rezidua[,i])
  u_med[i]=sum(rozdil_med[1:(n-1),i]
    *rozdil_med[2:n,i] < 0) + 1
  m_med[i]= sum(rozdil_med[,i] < 0)
  test_med[i]= (abs(u_med[i]-(m_med[i]+1)))
    /sqrt(m_med[i]*(m_med[i]-1)/(2*m_med[i]-1))
  rozhodnuti_med[i]=ifelse(test_med[i]>qnorm(0.975),
    "H0 zamítáme", "H0 nelze zamítnout")}
```