

UNIVERZITA PALACKÉHO V OLOMOUCI
PŘÍRODOVĚDECKÁ FAKULTA

DIPLOMOVÁ PRÁCE

Statistická analýza dat Energetického regulačního
úřadu



Katedra matematické analýzy a aplikací matematiky
Vedoucí diplomové práce: **doc. RNDr. Eva Fišerová, Ph.D**
Vypracovala: **Bc. Jana Prokopová**
Studijní program: N1103 Aplikovaná matematika
Studijní obor: Aplikace matematiky v ekonomii
Forma studia: prezenční
Rok odevzdání: 2015

BIBLIOGRAFICKÁ IDENTIFIKACE

Autor: Bc. Jana Prokopová

Název práce: Statistická analýza dat Energetického regulačního úřadu

Typ práce: Diplomová práce

Pracoviště: Katedra matematické analýzy a aplikací matematiky

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

Rok obhajoby práce: 2016

Abstrakt: Diplomová práce představí Energetický regulační úřad. Na datech, které sbírá tento úřad, předvede aplikaci spektrální analýzy a ARMA modelů. Ke zpracování dat je užit program R.

Klíčová slova: Energetický regulační úřad, statistika, časové řady, spektrální analýza, Boxova–Jenkinsova metodologie, predikce dat

Počet stran: 123

Počet příloh: 1

Jazyk: český

BIBLIOGRAPHICAL IDENTIFICATION

Author: Bc. Jana Prokopová

Title: The statistical analysis of the data of the Energy Regulatory Office

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: 2016

Abstract: The diploma thesis deals with Energy Regulatory Office. It presents the application of spectral analysis and ARMA models by using the data collected by this office. Software R is used to process obtained data.

Key words: Energy Regulatory Office, statistics, time series, spectral analysis, Box–Jenkins methodology, prediction

Number of pages: 123

Number of appendices: 1

Language: Czech

Prohlášení

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

V Olomouci dne

.....

podpis

Obsah

Použité značení	8
Úvod	9
1 Energetický regulační úřad	11
1.1 Zásadní body v průběhu vývoje ERÚ	14
2 Matematický aparát	17
2.1 Modelování časových řad	17
2.2 Predikce	18
2.3 Ověření kvality modelu	19
2.4 Spektrální analýza	21
2.4.1 Trendová složka	21
2.4.2 Určení modelu skrytých period	23
2.4.3 Periodogram a Fisherův test periodicity	24
2.5 Predikce ve spektrální analýze	26
2.6 Boxova–Jenkinsova metodologie	27
2.6.1 Lineární proces	31
2.7 Proces klouzavých součtů	33
2.8 Autoregresní proces	35
2.9 Smíšené procesy	37
2.10 Integrované smíšené modely	38
2.11 Sezónní integrované smíšené modely	40
2.12 Výstavba modelů	41
2.12.1 Identifikace modelu	42
2.12.2 Odhad parametrů modelu	49
2.12.3 Ověřování modelu	52
2.13 Předpovědi	56
2.13.1 Konstrukce předpovědí	58
3 Praktická část	60
3.1 Příprava dat, přípravné práce analytika	60
3.2 Aplikace spektrální analýzy	62
3.2.1 Očištění dat od trendu	62
3.2.2 Spektrální analýza	71
3.2.3 Spektrální analýza dat dovozu elektrické energie	71
3.2.4 Spektrální analýza dat ztrát elektrické energie v sítích	81
3.2.5 Spektrální analýza dat výroby elektrické energie	85
3.3 Aplikace Boxovy–Jenkinsovy metodologie	89
3.3.1 Boxova–Jenkinsova metodologie pro data dovozu	92
3.3.2 Boxova–Jenkinsova metodologie pro data ztrát	104

3.3.3	Boxova–Jenkinsova metodologie pro data výroby	109
3.4	Shrnutí výsledků	113
	Závěr	117
	Seznam obrázků	119
	Seznam tabulek	121
	Literatura	123

Poděkování

Ráda bych na tomto místě poděkovala vedoucí diplomové práce paní doc. RNDr. Evě Fišerové, Ph.D. za spolupráci, rady i za čas, který mi věnovala při konzultacích. Dále bych ráda poděkovala svým rodičům a příteli, kteří při mně stáli a podporovali mne v bádání spojeném s psaním diplomové práce.

Použité značení

\mathbb{R}	množina reálných čísel
t	index času
ω	úhlová frekvence
$\varepsilon_t \sim \text{WN}(0, \sigma_\varepsilon^2)$...	bílý šum, kde $E(\varepsilon_t) = 0$, $\text{var}(\varepsilon_t) = \sigma_\varepsilon^2$, $\text{cov}(\varepsilon_t, \varepsilon_s) = 0$, $t \neq s$
H_0	nulová hypotéza
γ_k	autokovarianční funkce v bodě k
ρ_k	autokorelační funkce v bodě k
ρ_{kk}	parciální autokorelační funkce
B	tzv. operátor zpětného posunutí
$\theta(B)$	operátor klouzavých součtů řádu q
$\varphi(B)$	autoregresní operátor řádu p
Δ	diferenční operátor
$\nu(B)$	zobecněný autoregresní operátor
L	délka periody v sezónních modelech
$\Phi(B^L)$	sezónní autoregresní operátor řádu P
$\Theta(B^L)$	sezónní operátor klouzavých součtů řádu Q
Δ_L	sezónní diferenční operátor
d	řád difference procesu y_t
D	řád sezónní difference procesu y_t

Úvod

Energetický regulační úřad vznikl jako ústřední orgán státní správy. Jeho hlavním úkolem je dohlížet na správnou funkci energetického trhu a kontrolovat dodržování zákona č. 458/2000 Sb., o podmínkách podnikání a o výkonu státní správy v energetických odvětvích a o změně některých zákonů. Téměř každý občan našeho státu o energetickém regulačním úřadu někdy slyšel. Bylo to v důsledku různé medializace tohoto úřadu. Nicméně jen malá část naší společnosti ví, proč úřad funguje, zná jeho členění či jeho historický vývoj. Na tyto otázky nalezneme odpověď v první kapitole této práce.

Vůči tomuto orgánu mají zainteresované subjekty povinnost poskytovat určité informace. Energetický regulační úřad informace shromažďuje a zveřejňuje je na oficiálních internetových stránkách. Jde o informace týkající se vývoje různých energetických ukazatelů v čase. Vybrané ukazatele analyzujeme v této diplomové práci. Na data nahlížíme jako na časovou řadu, neboli analyzujeme vývoj konkrétních pozorování v čase. Díky jejich analýze dojde k lepšímu pochopení různých zákonitostí, které v těchto časových řadách působí. Tyto zákonitosti se pokusíme nalézt s využitím spektrální analýzy a následně pomocí *ARMA* modelů.

Získané modely lze užít k predikci budoucího vývoje. Tato predikce je v blízké budoucnosti docela přesná, bohužel ale s rostoucí délkou se predikce stává více a více nepřesnou. Je tomu proto, že na každou časovou řadu působí mnoho faktorů. Vliv může mít změna legislativy, změna v užívané technologii, příchod nového subjektu na trh a mnoho dalších. Právě toto mění charakter průběhu časové řady a ovlivňuje kvalitu budoucí predikce.

Cílem této práce je odhalit zákonitosti působící ve zkoumaných datech, tyto zákonitosti co nejlépe popsat a zahrnout do vytvářeného modelu a následně zhodnotit, do jaké míry se nám analýza povedla. Po vytvoření modelu se pokusíme stanovit bodové i intervalové predikce budoucího vývoje. K celé této problematice potřebujeme nejdříve vysvětlit námi používaný matematický aparát, který je blíže popsán ve druhé kapitole. Vysvětlíme si pojem časová řada, popíšeme potřebné kroky k odhalení významných period v časové řadě či základní cha-

rakteristiku *ARMA* modelů. Třetí kapitola pojednává o zpracování zkoumaných dat s využitím programu *R*. Data si v ní připravíme ke zpracování, testujeme je pomocí Fisherova testu a vytváříme predikci z nalezených modelů. Nakonec na data aplikujeme Boxovu–Jenkinsovu metodologii. Dosažené výsledky praktické části jsou zobrazovány na přehledných grafech. V poslední kapitole této diplomové práce jsou dosažené výsledky zhodnoceny.

1. Energetický regulační úřad

Energetický regulační úřad (ERÚ) [9] zahájil svoji činnost 1. 1. 2001 podle zákona č. 458/2000 Sb., o podmínkách podnikání a o výkonu státní správy v energetických odvětvích a o změně některých zákonů (dále energetický zákon). Byl zřízen jako nový ústřední orgán státní správy se sídlem v Jihlavě. Jeho dislokovaná pracoviště se nacházejí v Praze a Ostravě. Od 1. 8. 2011 úřad řídí předsedkyně Ing. Alena Vitásková, která je volena na dobu šesti let. Na návrh vlády ji jmenuje prezident republiky. Paní předsedkyni zastupuje pět místopředsedů.

Mezi hlavní činnosti ERÚ patří

- vykonávání dohledu nad trhy v energetických odvětvích,
- ochrana zájmů zákazníků, spotřebitelů a oprávněných požadavků držitelů licencí,
- regulování cen energií,
- podporování využívání obnovitelných a druhotných zdrojů energie,
- vykonávání dohledu nad soutěžními podmínkami v energetických odvětvích,
- podporování hospodářské soutěže v energetických odvětvích.

Energetický regulační úřad má přesně danou organizační strukturu. Jednotlivé části této struktury a jejich náplň práce jsou popsány v následujících bodech.

1. *Úsek 1. místopředsedy* je zodpovědný za rozhodnutí ředitelů sekce kontroly a sekce provozní. Pozici prvního místopředsedy zastává RNDr. Otto Gold.
2. *Oddělení interního auditu* se stará o dodržování zákona č. 320/2001 Sb., o finanční kontrole ve veřejné správě a je podřízeno přímo předsedkyni úřadu.
3. *Bezpečnostní ředitel* zajišťuje veškeré činnosti související se zvládáním krizových situací. Mezi jeho úkoly patří také zabezpečení ochrany utajovaných

informací a kybernetické bezpečnosti úřadu. Ředitel se zodpovídá přímo předsedkyni ERÚ.

4. *Odbor kanceláře předsedkyně* zabezpečuje činnosti související s řízením úřadu, plánováním a poskytováním informací ostatním institucím. Odbor řídí Ing. Svatopluk Srba.
5. *Sekce regulace* rozhoduje různé spory, jako např. mezi držiteli licencí nebo mezi držiteli licencí a jejich zákazníky. V její kompetenci je příprava cenových rozhodnutí o energiích, schvalování pravidel pro provoz soustav v plynárenství a teplárenství. Vytvářejí se zde také vyhlášky a pravidla pro správné fungování trhu. Pan Ing. Jan Nehoda je místopředsedou ERÚ a ředitel této sekce.
6. *Sekce legislativně správní* kontroluje a vykonává dozor nad ochranou spotřebitele. Zabezpečuje také legislativně správní činnost. V jejím čele stojí dvě osobnosti. Jednou z nich je ředitelka Mgr. Irena Dóményová, LL.M. a místopředsedkyně JUDr. Renata Vesecká, Ph.D.
7. *Odbor podporovaných zdrojů* se řídí zákonem č. 165/2012 Sb. o podporovaných zdrojích energie a realizuje dozor nad dodržováním energetického zákona a zákona o cenách. Při výkonu kontroly se postupuje dle zákona č. 552/1991 Sb., o státní kontrole. Pan Ing. Martin Laštůvka se stal místopředsedou úřadu a ředitelem této sekce.
8. *Sekce provozní* spravuje samostatnou kapitolu státního rozpočtu dle § 17 energetického zákona. Mezi její další kompetence patří správa majetku, ekonomická činnost, mzdová, personální, organizační a provozní činnost. Jejím ředitelem je Ing. Vladimír Kabelka, MBA.
9. *Sekce kontroly* je jednou z nejdůležitějších částí ERÚ. V této sekci se rozhoduje o udělení, změně nebo zrušení licence podnikatelů v energetických odvětvích. Sekce také dozoruje nad dodržováním energetického zákona a zákona o cenách podle zákona č. 552/1991 Sb., o státní kontrole.

10. *Sekce vnějších vztahů* zajišťuje komunikaci úřadu s veřejností, s ostatními orgány státní správy a se zahraničními institucemi. V její kompetenci je zajištění legislativních a organizačních náležitostí pro vznik Energetického ombudsmana. Tiskovým mluvčím ERÚ je Jiří Chvojka, B.A., který zároveň řídí tuto sekci.
11. *Sekce pro evropské záležitosti a strategii* je poměrně novou sekcí. Byla vytvořena z důvodu účinnějšího propojení zahraničních struktur ACER, CEER a uplatňování nařízení REMIT. Ředitelka této sekce paní Ing. Martina Krčová, MBA je také místopředsedkyní ERÚ.

Agentura pro spolupráci energetických regulačních orgánů (ACER) zahájila svoji činnost 3. 3. 2011 na základě nařízení Evropského parlamentu a Rady EU č. 713/2009. Působí jako koordinátor spolupráce energetických regulačních orgánů, snaží se o vytvoření jednotného trhu s energií a společných provozních a obchodních pravidel. Kontroluje práci evropských provozovatelů přenosových a přepravních soustav, rozvoj sítí a funkčnost trhů s elektřinou a plynem.

Rada evropských energetických regulátorů (CEER) je dobrovolné sdružení energetických regulačních úřadů zemí Evropské unie. Pomáhá vytvářet jednotný, udržitelný a konkurenceschopný energetický trh, jednotné podmínky pro všechny jeho účastníky a jednotnou evropskou energetickou legislativu.

REMIT je nařízení Evropského parlamentu a Rady EU č. 1227/2011 o integritě a transparentnosti velkoobchodního trhu s energií, které vstoupilo v platnost 28. 12. 2011. Toto nařízení zavádí konzistentní rámec pro celou Evropskou unii. Definiuje manipulaci s trhem a obchodování zasvěcených osob na velkoobchodních trzích s energií. Zakazuje zneužívání trhu a určuje tresty při porušení tohoto nařízení. Vymezuje monitorovací rámec velkoobchodního trhu s energií a nařizuje účastníkům trhu zveřejňovat zásadní informace. Nařízení zavazuje úřad ke spolupráci s agenturou ACER a s vnitrostátními regulačními orgány při kontrole trhů a vyšetřování porušení závazných nařízení. ERÚ na základě nařízení REMIT vytvořilo národní registr účastníků trhu.

Počet zaměstnanců ERÚ se v průběhu let měnil. Při zahájení činnosti úřadu

byl počet pouhých 50 zaměstnanců. K 31.12.2014 úřad zaměstnával 224 lidí. Plánovaný počet zaměstnanců pro rok 2015 byl 254.

1.1. Zásadní body v průběhu vývoje ERÚ

Energetický regulační úřad [9] zahájil svoji činnost 1.1.2001. Je zřízen a řídí se dle energetického zákona. Jde o správní úřad pro výkon regulace v energetice, jehož primární činností je dohlížet na kvalitu a spolehlivost dodávek energií spotřebitelům, chránit jejich zájmy, podporovat konkurenci a stabilitu cenové úrovně. Energetický regulační úřad asistoval při přípravách ke vstupu České republiky do Evropské unie. V tomto období byl vytvořen funkční regulační rámec, který podporoval liberalizaci elektroenergetiky a plynárenství. Úřad také dokončoval práce na prováděcích předpisech k energetickému zákonu. V roce 2001 působili postupně na úřadu tři předsedové, paní Ing. Jana Novotná v období od 1. 1. 2001 do 7. 2. 2001. Pan Ing. Pavel Prouza její funkci převzal 8. 2. 2001 a byl pověřen řízením do 21. 3. 2001. Následně byl zvolen předsedou Ing. Pavel Brychta, CSc., který zde působil až do 31. 8. 2004.

V roce 2002 byla vytvořena nová struktura výpočtu ceny elektřiny, začala etapa otevírání trhu s elektřinou a zahájily se přípravné práce pro vznik pravidel fungujícího trhu s plynem.

Rok 2003 byl rokem vstupu úřadu do Rady evropských energetických regulátorů. Změnila se legislativa EU a úřad se musel přizpůsobit upraveným podmínkám a organizaci vnitřních energetických trhů včetně způsobu podpory využívání obnovitelných zdrojů energie a přeshraniční výměny v rámci jednotného trhu unie.

Po přípravných operacích se Česká republika v roce 2004 stala členem EU. Tímto vstupem se pro nás stala pravidla pro vnitřní trh s elektřinou závazná.

V roce 2005 pokračovala liberalizace na trhu s elektřinou a bylo zahájeno otevírání trhu s plynem. ERÚ pomáhal zdokonalovat legislativu v návaznosti na zákon o podpoře využívání obnovitelných zdrojů. V jeho spolupráci s ministerstvem průmyslu a obchodu vznikla první Národní zpráva České republiky

o elektroenergetice a plynárenství za předešlý rok. Změnil se také předseda ERÚ, 1. 9. 2004 se jím stal Ing. Josef Fírt. Ten funkci zastával do 4. 7. 2011.

Od 1. 1. 2006 si všichni odběratelé elektřiny mohli vybrat svého dodavatele elektrické energie. Úřad v tomto roce přestal určovat konečnou cenu elektřiny, ale stanovil pouze její regulované složky. Stal se také uznávaným orgánem při řešení sporů mezi dodavatelem energií a jejich odběrateli a držiteli licencí.

Všichni zákazníci v oblasti plynárenství si mohou od 1. 1. 2007 zvolit svého dodavatele. V tomto roce je také připravována nová evropská legislativa v oblasti energetiky. ERÚ je pověřen k vytvoření několika nařízení z legislativního balíčku v oblasti energetiky.

V roce 2008 byla dokončena novela energetického zákona č. 158/2009 Sb., která přinesla mnohé změny. Patří mezi ně upřesnění ustanovení upravující působnost ERÚ při řešení sporů. Důraz je kladen na zvýšení bezpečnosti, spolehlivosti, kvality dodávek a na spolupráci provozovatelů přenosových soustav se zahraničím. Úřad také pokračuje v mezinárodní spolupráci v rámci EU.

Již zmíněná novela vstoupila v platnost 4. 7. 2009. Energetický regulační úřad vypracoval novelizaci vyhlášek z oblasti regulace. Přichystal také metodiku regulace síťových činností v odvětví elektroenergetiky a plynárenství pro následující pětileté období. V tomto roce narostl počet výrobců elektřiny z obnovitelných zdrojů energií, a proto došlo ke zpracování novely zákona o podpoře využívání obnovitelných zdrojů.

Pro rok 2010 je charakteristická tvorba různých novel. Šlo o přípravu novely energetického zákona, zákona o podpoře využívání obnovitelných zdrojů, novela vyhlášky o podmínkách připojení k elektrizační soustavě, novela vyhlášky o Pravidlech trhu s plynem a Pravidlech trhu s elektřinou. Pro následující rok byly stanoveny základní principy regulace cen v energetice.

V roce 2011 se rozšířily pravomoci úřadu. ERÚ získal větší pravomoci v oblasti dozoru nad fungováním trhu s elektřinou a plynem, ochrany práv zákazníků při sjednávání smluv a kontrolních pravomocí. Byl zahájen projekt propojení maďarské, české a slovenské tržní oblasti.

V roce *2012* vzniklo další dislokované pracoviště úřadu. Toto pracoviště zahájilo svoji činnost v Ostravě. ERÚ klade zvýšený důraz na kvalitu dodávek energií, rovnost práv na trhu a etiku prodeje. Díky novele energetického zákona rozšířil úřad spektrum svých pravomocí, zejména v kontrole energetického trhu. ERÚ spolupracoval také na přípravě zákona o podporovaných zdrojích energií. V tomto roce vznikl Etický kodex obchodníka v energetických odvětvích, který upravuje základní pravidla etického a profesionálního jednání obchodníka s elektřinou a plynem. Úřad se tedy více soustředí na ochranu práv spotřebitelů a zákazníků.

Novelizace zákonů byly připravovány i v roce *2013*. Šlo o novelu energetického zákona, zákona o podporovaných zdrojích energií, návrh zákona o integritě a transparentnosti velkoobchodních trhů s energií. Je vyvinuta snaha o větší spolupráci na energetickém trhu.

V roce *2014* se úřad soustředil na mezinárodní spolupráci a zkvalitnění legislativy v oblasti energetiky. Vznikla sekce pro evropské záležitosti a strategie. V České republice se konala mezinárodní konference REMIT a prohlubovala se mezinárodní spolupráce v oblasti regionálních aktivit. Připravovala se novela energetického zákona a zákona o podporovaných zdrojích energie.

2. Matematický aparát

Cílem této diplomové práce je analýza časových řad a následná predikce budoucího vývoje. K tomu je potřeba znalost určité tematiky, a ta je uvedena v následujících podkapitolách. V nich čerpáme převážně z [2], [4] a [5].

2.1. Modelování časových řad

Časovou řadou chápeme posloupnost jednoznačně uspořádaných dat od minulosti po přítomnost, která jsou věcně a prostorově srovnatelná. Tyto řady mají různé členění, např.:

- a) intervalové - řady, jejichž ukazatel závisí na délce sledovaného období
- b) okamžikové - představují ukazatele vztahující se k určitému dni
- a) ekvidistantní - ukazatelé mají stejně dlouhé intervaly mezi pozorováními
- b) neekvidistantní - intervaly mezi sledovanými ukazateli nejsou stejně dlouhé.
- a) dlouhodobé - časové období mezi dvěma ukazateli je rok a více
- b) krátkodobé - časové rozpětí mezi dvěma ukazateli je menší než rok, např. měsíc, týden

Dekompozice časových řad se dle [5] provádí rozčleněním časové řady na složku trendovou (T_t), cyklickou (C_t), sezónní (S_t) a náhodnou (ε_t). Index t je časovou doménou, kde $t = 1, 2, \dots, n$. Časovou řadu můžeme dekomponovat buď aditivně, rozklad řady je ve tvaru

$$y_t = T_t + C_t + S_t + \varepsilon_t,$$

kde jednotlivé složky uvažujeme ve skutečných absolutních hodnotách a měření probíhá v jednotkách řady y_t . Tento typ dekompozice se využívá, pokud variabilita hodnot časové řady je přibližně konstantní v čase. Nebo můžeme zvolit multiplikativní přístup, kdy rozklad řady zapíšeme jako

$$y_t = T_t C_t S_t \varepsilon_t.$$

V multiplikatívni dekompozici je trendová složka v absolutní hodnotě a je měřena ve stejných jednotkách jako y_t . Cyklická, sezónní a náhodná složka jsou bezrozměrné v relativních hodnotách. Multiplikatívni dekompozici využijeme v případě, že variabilita hodnot časové řady se v čase mění a nebo roste.

Trendem časové řady chápeme hlavní tendenci dlouhodobého vývoje sledovaných ukazatelů v průběhu času. *Periodická složka* zahrnuje cyklickou a sezónní složku. Cyklická složka popisuje dlouhodobé odchylky od trendu s periodou delší než roční. Sezónní složka charakterizuje pravidelně se opakující výkyvy od trendu s periodou kratší než roční. *Náhodná složka* je ta část časové řady, která zbývá po určení trendu a periodické složky. Snažíme se, aby byla co nejmenší, protože jsou to vlivy, které model vůbec nepopisuje. Při dekompozici časových řad zjistíme, že ne všechny řady obsahují všechny složky. Konkrétní řadu může charakterizovat jen jedna či dvě složky.

U neekvidistantních časových řad se pro vzájemnou porovnatelnost ukazatelů provádí očišťování ukazatelů od kalendářních variací:

$$y_t^{(0)} = y_t \frac{\bar{d}}{d_m}, \quad t = 1, 2, \dots, n,$$

kde $y_t^{(0)}$ jsou očištěné hodnoty ukazatelů, y_t jsou hodnoty jednotlivých ukazatelů, \bar{d} průměrný počet kalendářních dní, d_m je počet dní v m -tém kalendářním měsíci a $m = 1, 2, \dots, 12$.

V průběhu naší analýzy se snažíme být co nejflexibilnější, a proto používáme modelování časových řad pomocí *Spektrální analýzy* a *Boxovy–Jenkinsovy metodologie*.

2.2. Predikce

Jedním z důvodů, proč se provádí analýza časových řad, je snaha z nalezeného modelu sestavit předpověď budoucího chování zkoumané časové řady.

Bodová předpověď je odhadnutá budoucí hodnota sledované časové řady. Bohužel právě proto, že jde o odhad budoucí hodnoty, je bodová předpověď zatížena určitou chybou.

Předpovědní interval udává dolní a horní hranici, mezi kterou bude ležet příslušná budoucí hodnota zkoumaných dat s danou dostatečně velkou pravděpodobností.

Chyba předpovědi představuje rozdíl mezi predikovanými hodnotami a reálnými hodnotami. Tato chyba vzniká proto, že při vytváření předpovědi vycházíme ze známých hodnot a odhadneme z nich budoucí hodnoty. Nikdy nemůžeme určit přesné hodnoty nastávající v reálném čase, ale jen jejich odhady. Je tomu tak z důvodu vnějších vlivů, které odhadnutý model nemůže zachytit. Tuto chybu můžeme určit až v době, kdy odhadnuté hodnoty mají svůj reálný protějšek. Chybu předpovědi značíme ϵ_t a určíme ji ze vztahu

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

kde y_t jsou skutečně naměřené hodnoty a \hat{y}_t jsou příslušné odhadnuté hodnoty.

2.3. Ověření kvality modelu

Naším úkolem je analyzovat reálná data pomocí modelování časových řad, a proto bychom si měli představit nějaké hodnotitele kvality určených modelů.

Index determinace, který značíme R^2 , je jedním z ukazatelů vhodnosti aplikovaného modelu na sledovaná data. Může nabývat hodnot v intervalu $\langle 0, 1 \rangle \in \mathbb{R}$. Čím více se R^2 blíží 1, tím lépe jsme vysvětlili a popsali zkoumaná data a tím kvalitněji jsme určili hledanou regresní funkci. Naopak, pokud je index blízký 0, model není vhodný. Index determinace definuje vzorec [5]

$$R^2 = 1 - \frac{\sum_{t=1}^n (y_t - \hat{y}_t)^2}{\sum_{t=1}^n (y_t - \bar{y})^2},$$

kde y_t jsou původní data, \hat{y}_t představují data odhadnutá modelem a \bar{y} je aritmetický průměr původních dat. Velkou nevýhodou indexu determinace je jeho závislost na počtu parametrů modelu. Čím více má model parametrů tím vyšší dostáváme hodnotu indexu. Často se tedy využívá pro hodnocení kvality získaného modelu *modifikovaný index determinace* R_M^2 , který tento nedostatek od-

straňuje. Je dán vztahem [4]

$$R_M^2 = R^2 - \frac{(1 - R^2)(k - 1)}{n - k},$$

kde n je počet pozorování a k je počet parametrů modelu.

Jiným ukazatelem míry kvality modelu je *střední čtvercová chyba* MSE . Pomocí vzorce [4]

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

můžeme porovnávat nalezené modely mezi sebou. Nejlepším modelem označíme ten, který má minimální hodnotou MSE . Nevýhodou tohoto ukazatele je jednotka, ve které je vyjádřen. Ze vzorce je patrné, že se jedná o druhou mocninu původních jednotek hodnot časové řady. Tuto negativní vlastnost odstraníme užitím $RMSE$, které je definováno vztahem [6]

$$RMSE = \sqrt{MSE}.$$

Pro vzájemné porovnávání časových řad potřebujeme bezrozměrný ukazatel. Získáme jej normováním hodnoty $RMSE$. Normovaný ukazatel značíme $NRMSE$ a je dán vztahem

$$NRMSE = \frac{RMSE}{y_{t_{max}} - y_{t_{min}}}. \quad (1)$$

Označení $y_{t_{max}}$ používáme pro maximální hodnotu časové řady y_t a obdobně $y_{t_{min}}$ označuje minimální prvek řady y_t . K porovnávání modelů můžeme také využít *součtu čtvercových chyb* SSE , jehož hodnotu určíme vzorcem [5]

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

Jako nejlepší zvolíme model s nejmenší hodnotou SSE . *Střední absolutní chyba* MAE je ve tvaru [6]

$$MAE = \frac{1}{h} \sum_{t=n+1}^{n+h} |y_t - \hat{y}_t|,$$

kde h je délka předpovědi. Používá se pro hodnocení předpovědních chyb v časové řadě. Nejlepší je model s nejmenší hodnotou MAE . Ukazatelem kvality předpovědi je také *střední absolutní procentuální chyba* $MAPE$ ve tvaru [6]

$$MAPE = \frac{100}{h} \sum_{t=n+1}^{n+h} \left| \frac{y_t - \hat{y}_t}{y_t} \right|.$$

Tento ukazatel je bezrozměrný, můžeme jej užívat k vzájemnému porovnávání kvality předpovědí časových řad. Je vyjádřen v procentech a nejvhodnější predikce modelu je ta s nejnižší hodnotou $MAPE$.

2.4. Spektrální analýza

Tento přístup zkoumání časových řad se aplikuje na data, která obsahují sezónní a cyklickou složku. Časové řady se při užití spektrální analýzy popisují jako směs sinusovek a kosinusovek s různou frekvencí a amplitudou. Při využití spektrální analýzy vlastně hledáme významné složky periodicity v časových řadách, a nebo porovnáváme chování několika periodických časových řad navzájem.

2.4.1. Trendová složka

Časová řada, kterou chceme zkoumat, může obsahovat i trendovou složku. Při aplikaci spektrální analýzy modelujeme pouze periodickou složku. Je tedy vhodné prozkoumat, zda časová řada neobsahuje nějaký trend. Základní druhy trendových křivek jsou

- *konstantní trend*

$$T_t = \beta_0,$$

- *lineární trend*

$$T_t = \beta_0 + \beta_1 t,$$

- *kvadratický trend*

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

- *polynomický trend řádu r*

$$T_t = \beta_0 + \beta_1 t + \dots + \beta_r t^r,$$

- *exponenciální trend*

$$T_t = \alpha \beta^t, \quad \beta > 0,$$

- *modifikovaný exponenciální trend*

$$T_t = \gamma + \alpha \beta^t, \quad \beta > 0,$$

- *logistický trend*

$$T_t = \frac{\gamma}{1 + \alpha \beta^t}, \quad \beta > 0, \gamma > 0,$$

- *Gompertzova křivka*

$$\log T_t = \gamma + \alpha \beta^t, \quad \beta > 0,$$

kde $t = 1, \dots, n$. Uvedené funkce vycházejí z aplikace teorie lineárních regresních modelů a optimalizace pomocí metody nejmenších čtverců. Jejich bližší popis včetně odvození jednotlivých vzorců lze nalézt v literatuře [5, str. 30].

Pokud data trend obsahují, pak můžeme využít pro jejich zachycení aditivní dekompozici tvaru $y_t = T_t + \varepsilon_t$, nebo multiplikativní dekompozici $y_t = T_t \varepsilon_t$. Trend musíme odhalit a následně časovou řadu od tohoto trendu očistit. Pro určování nejvhodnějšího tvaru trendové funkce volíme postupné prokládání dat trendovými funkcemi. O vhodnosti trendové funkce budeme rozhodovat pouze vizuálně za využití grafů. Vybereme ten z trendů, který bude nejlépe kopírovat chování sledovaných dat. Nebudeme využívat index determinace, protože data neobsahují pouze trendovou složku. Jeho hodnota by byla příliš zkreslená ostatními složkami časové řady a neměla by vhodné vypovídající schopnosti.

Po určení trendové složky musíme naše data od tohoto trendu očistit. Model s aditivní dekompozicí očistíme od trendové složky funkcí $y_t^{(0)} = y_t - T_t$. V případě multiplikativní dekompozice vznikne očištěná časová řada užitím $y_t^{(0)} = \frac{y_t}{T_t}$. Dále budeme zkoumat vlastnosti nové časové řady $y_t^{(0)}$, kterou opět označíme jako y_t .

2.4.2. Určení modelu skrytých period

V následujícím textu budeme užívat pojmy frekvence a perioda, a proto si je musíme vysvětlit. *Úhlová frekvence* ω , popsaná např. v [5], je dána v radiánech za uvažovanou časovou jednotku, kterou tvoří časový interval mezi dvěma sousedními hodnotami časové řady. Například frekvence 2π radiánů za časovou jednotku znamená uskutečnění právě jednoho cyklu dané funkce za uvažovanou časovou jednotku. Pro úhlovou frekvenci platí, že čím je frekvence větší, tím častěji se v jejím průběhu střídají jednotlivé cykly. Vypočítáme ji jako $\omega_j = \frac{2\pi j}{n}$, kde $0 < j \leq n$. Její převrácenou hodnotu $\frac{2\pi}{\omega}$ nazýváme *délkou periody*, což je doba, za kterou proběhne jeden cyklus.

Ve spektrální analýze se snažíme časové řady zachytit modelem skrytých period. Model je definován funkcí

$$y_t = \mu + \sum_{j=1}^p (\alpha_j \cos(\omega_j t) + \beta_j \sin(\omega_j t)) + \varepsilon_t, \quad t = 1, \dots, n, \quad (2)$$

kterou jsme převzali z [5, str. 222]. Řada obsahuje různé frekvence $\omega_1, \dots, \omega_p$ z intervalu $(0, \pi)$ a neznámé parametry μ, α_j, β_j . Člen ε_t nazýváme bílý šum (white noise), pro který platí $\varepsilon_t \sim WN(0, \sigma_\varepsilon^2)$, tedy jsou splněny předpoklady normality $E(\varepsilon_t) = 0$, $var(\varepsilon_t) = \sigma_\varepsilon^2$ a dále platí $cov(\varepsilon_t, \varepsilon_s) = 0$, kde $t \neq s$ a rozptyl $\sigma_\varepsilon^2 > 0$ je neznámý parametr modelu. Odhady neznámých parametrů určíme pomocí metody nejmenších čtverců následovně [5, str. 224]

$$\hat{\mu} = \bar{y} = \frac{1}{n} \sum_{t=1}^n y_t, \quad (3)$$

$$\hat{\alpha}_j = \frac{2}{n} \sum_{t=1}^n y_t \cos(\omega_j t), \quad (4)$$

$$\hat{\beta}_j = \frac{2}{n} \sum_{t=1}^n y_t \sin(\omega_j t). \quad (5)$$

Odhady lze vypočítat pouze v případě, když známe významné periody, které určíme pomocí periodogramu a Fisherova testu periodicity.

2.4.3. Periodogram a Fisherův test periodicity

Nyní budeme zkoumat významnost periodické složky, která se skládá ze složky sezónní a cyklické. Nejdříve musíme vypočítat hodnoty periodogramu a ty následně dosadíme do Fisherova testu. Pro odhalení významných period budeme užívat Fisherův test periodicity. Ověříme významnost jednotlivých period a pouze významné poté zařadíme do výsledného modelu skrytých period. Celý postup si nyní popíšeme podrobněji.

Hodnoty *periodogramu* získáme dosazením očištěných dat y_t do následujícího vzorce

$$I(\omega_j) = \frac{1}{4\pi} \left[\frac{2}{n} \left(\sum_{t=1}^n y_t \cos(\omega_j t) \right)^2 + \frac{2}{n} \left(\sum_{t=1}^n y_t \sin(\omega_j t) \right)^2 \right], \quad (6)$$

kde parametr $t = 1, \dots, n$ je čas a n představuje počet pozorování [5]. Úhlovou frekvenci ω_j vypočítáme ze vztahu $\omega_j = \frac{2\pi j}{n}$, kde $j = 1, \dots, m$. Číslo m vypočítáme jako celou část z čísla $\frac{(n-1)}{2}$.

Získané hodnoty periodogramu využijeme ve *Fisherově testu periodicity* [5]. Takto budeme určovat významné hodnoty periodogramu a zanedbáme ty nevýznamné. Testujeme hypotézu

$$H_0 : y_t = a_0 + \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma_\varepsilon^2),$$

tzn. za platnosti hypotézy H_0 : v modelu neexistují žádné významné periody, proti alternativní hypotéze H_1 : v modelu se nachází alespoň jedna významná perioda. Platí tedy, kolikrát zamítneme nulovou hypotézu, tolik nám test nalezne významných period.

Podrobný postup Fisherova testu periodicity je následující. Nejdříve vypočítáme hodnoty periodogramu dle (6) pro frekvence $\omega_j = \frac{2\pi j}{n}$, kde $j = 1, \dots, m$. Poté vypočítáme normované hodnoty

$$V_j = \frac{I(\omega_j)}{\sum_{i=1}^m I(\omega_i)}, \quad (7)$$

kde $j = 1, \dots, m$. Normování se provádí k odstranění závislosti rozdělení dat na neznámém parametru modelu $\sigma_\varepsilon^2 > 0$. Následně určíme testovou statistiku

$$W = \max_{j=1, \dots, m} V_j. \quad (8)$$

Periodu označíme za významnou, pokud platí vztah

$$W > g_F. \quad (9)$$

Hodnota g_F označuje kritickou hodnotu Fisherova testu na zvolené hladině významnosti. Tyto hodnoty nalezneme v tabulkách [13, str. 12]. Pro získání kritických hodnot můžeme využít také vzorec z [5]

$$P(W > x) = m(1-x)^{m-1} - \binom{m}{2}(1-2x)^{m-1} + \binom{m}{3}(1-3x)^{m-1} - \dots,$$

kde $x \in (0, 1)$ a výpočet probíhá do zachování kladnosti členů $(1-jx)$. Pokud platí (9), zamítáme nulovou hypotézu a danou periodu označíme za významnou. V testování budeme pokračovat stejným způsobem, pouze z testovaných hodnot vypustíme současnou maximální hodnotu periodogramu. Také hodnotu m nahradíme $m-1$.

Časovou řadou se složenou periodicitou je nazývaná taková časová řada, která obsahuje dvě a více významných period s různou frekvencí. Fisherův test nedává moc kvalitní výsledky při složené periodicitě, proto byly navrženy různé modifikace. Jako příklad modifikace tohoto testu si uvedeme *Siegelovu modifikaci*. Používáme v ní testovou statistiku z [5], která je daná vztahem

$$T_\lambda = \sum_{j=1}^m (V_j - \lambda g_F)^+, \quad (10)$$

kde $(V_j)^+$ značí kladnou část čísla V_j a určíme ji jako $(V_j)^+ = \max(V_j, 0)$. Konstantu λ si zvolíme dopředu z intervalu $(0, 1)$, doporučuje se hodnota 0.6. Kritickou hodnotu g_F Fisherova testu na zvolené hladině významnosti opět získáme z tabulek [13, str. 12]. Pokud je $T_\lambda > t_\lambda$, pak zamítáme nulovou hypotézu. Kritické hodnoty t_λ na dané hladině významnosti jsou pro různá λ uvedeny taktéž v tabulkách [13, str. 12].

2.5. Predikce ve spektrální analýze

Záměrem téměř každé analýzy dat je nalezení modelu, kterým lze analyzovaná data popsat a ze kterého lze predikovat další vývoj dat. Ani zde tomu nebude jinak a po správném určení modelu vytvoříme predikci budoucího chování analyzované časové řady. K predikci nám poslouží model

$$y_t = T_t + \mu + \sum_{j=1}^p (\alpha_j \cos(t\omega_j) + \beta_j \sin(t\omega_j)) + \varepsilon_t,$$

kde $t = 1, \dots, n$ je čas, T_t představuje trendovou složku časové řady. Časovou řadu jsme očistili od trendové složky, následně jsme užitím vztahů (3), (4) a (5) vypočítali neznámé odhady parametrů μ, α_j, β_j . Úhlové frekvence $\omega_1, \dots, \omega_p$ jsou z intervalu $(0, \pi)$ a ε_t je bílý šum, který splňuje $E(\varepsilon_t) = 0$, $\text{var}(\varepsilon_t) = \sigma_\varepsilon^2$, $\text{cov}(\varepsilon_t, \varepsilon_s) = 0$, kde $t \neq s$, $\sigma_\varepsilon^2 > 0$ je opět neznámý parametr.

Pro budoucí pozorování v čase $n + h$ je předpovědní interval na hladině spolehlivosti $1 - \alpha$ ve tvaru [8]

$$\left(\hat{y}_{n+h} - t_{n-p-r}(\alpha) \sqrt{\hat{\sigma}_\varepsilon^2 (1 + x'_{n+h} (X'X)^{-1} x_{n+h})}; \right. \\ \left. \hat{y}_{n+h} + t_{n-p-r}(\alpha) \sqrt{\hat{\sigma}_\varepsilon^2 (1 + x'_{n+h} (X'X)^{-1} x_{n+h})} \right),$$

kde p je počet významných period v modelu a r je řád polynomické trendové funkce. Odhad \hat{y}_{n+h} představuje bodový odhad budoucích hodnot časové řady v čase n o h kroků dopředu, který určíme ze vztahu $\hat{y}_{n+h} = \mathbf{x}_{n+h} \mathbf{b}$, přičemž

$$\mathbf{x}_{n+h} = \begin{pmatrix} 1 \\ n+h \\ (n+h)^2 \\ \vdots \\ (n+h)^r \\ \cos((n+h)\omega_1) \\ \sin((n+h)\omega_1) \\ \vdots \\ \cos((n+h)\omega_p) \\ \sin((n+h)\omega_p) \end{pmatrix}'.$$

Vektor \mathbf{b} určíme ze vztahu $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$. Kritickou hodnotu studentova t-rozdělení o $n-p-r$ stupních volnosti na hladině významnosti α vyhledáme v tabulkách. Odhad parametru $\hat{\sigma}_\varepsilon^2$ vypočítáme vztahem

$$\hat{\sigma}_\varepsilon^2 = \sum_{t=1}^n \frac{(y_t - \hat{y}_t)^2}{n-p-r},$$

kde $\hat{y}_t = \mathbf{X}\mathbf{b}$. Designová matice \mathbf{X} pro model s polynomickým trendem řádu r je tvaru [8]

$$\mathbf{X} = \begin{pmatrix} 1 & t_1 & \dots & t_1^r & \cos(t_1\omega_1) & \sin(t_1\omega_1) & \dots & \cos(t_1\omega_p) & \sin(t_1\omega_p) \\ 1 & t_2 & \dots & t_2^r & \cos(t_2\omega_1) & \sin(t_2\omega_1) & \dots & \cos(t_2\omega_p) & \sin(t_2\omega_p) \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & t_n & \dots & t_n^r & \cos(t_n\omega_1) & \sin(t_n\omega_1) & \dots & \cos(t_n\omega_p) & \sin(t_n\omega_p) \end{pmatrix},$$

$n \times (r+1+2p)$

kde $\mathbf{t} = (t_1, t_2, t_3, \dots, t_n)' = (1, 2, 3, \dots, n)'$ je čas a $j = 1, \dots, p$ představuje počet významných period. Po dosazení všech hodnot získáme hledaný empirický předpovědní interval na hladině spolehlivosti $1 - \alpha$.

2.6. Boxova–Jenkinsova metodologie

Předmětem zkoumání v Boxově–Jenkinsově metodologii je náhodná složka. Náhodná složka může být tvořena i závislými náhodnými veličinami. Tento přístup k analýze časových řad se nezabývá jen náhodnou složkou, ale můžou se pomocí něho modelovat i časové řady s trendem a sezónností. Využívá se k tomu model *ARIMA*, který modeluje trendovou složku stochasticky. Model je popsán v kapitole 2.10. Pro modelování sezónní složky se užívá model *SARIMA*, který modeluje stochasticky sezónní i trendovou složku. Odpovídající teorie je uvedena v kapitole 2.11.

Boxova–Jenkinsova metodologie se využívá hlavně kvůli své schopnosti rychle se adaptovat na změny v charakteru časové řady tak, že dostáváme kvalitní výsledky a dobré předpovědi. Proces zkoumání časové řady se člení na tři fáze

1. identifikace, kde se snažíme určit vhodný typ modelu časové řady,

2. odhad parametrů modelu,
3. verifikace, která představuje ověření správnosti modelu.

Všechny metody určené ke zkoumání dat mají i své slabé stránky, ani zde tomu není jinak. Mezi negativa této metodologie můžeme zahrnout potřebu mít poměrně velké množství dat (alespoň 50, ale spíše více), abychom vytvořili spolehlivý model. Zpracovávání dat těmito modely je poměrně časově náročné. Posledním zmíněným negativem je obtížná interpretovatelnost zjištěných výsledků.

Nyní již přejdeme k vlastní terminologii. Musíme si uvést několik základních pojmů, abychom chápali základní souvislosti. Budeme čerpat z [5].

Stochastickou závislost veličin t a s měříme pomocí *kovariance*, která je dána vztahem

$$\text{cov}(t, s) = E(t - E(t))(s - E(s)) = E(ts) - E(t)E(s).$$

Funkce daná předpisem

$$\text{cor}(t, s) = \frac{\text{cov}(t, s)}{\sqrt{\text{var}(t)\text{var}(s)}}, \quad \text{var}(t) \neq 0, \text{var}(s) \neq 0,$$

vyjadřuje *korelační koeficient*, který nabývá hodnot $\text{cor}(t, s) \in \langle -1, 1 \rangle$.

Jeden ze základních předpokladů pro užití Boxovy–Jenkinsovy metodologie je předpoklad *stacionarity*. Pojem představuje ustálené pravděpodobnostní chování stochastického procesu. Rozlišujeme

1. *striktní stacionaritu*, která je silnější než slabá stacionarita. Pravděpodobnostní chování stochastického procesu je nezávislé na posunech v čase. Neboli náhodný vektor $(y_{t_1}, \dots, y_{t_n})$ má stejné pravděpodobnostní rozdělení jako je rozdělení vektoru $(y_{t_1+h}, \dots, y_{t_n+h})$ pro libovolné h .
2. *slabou stacionaritu*, která nastává pro stochastický proces s konstantní střední hodnotou, s konstantním rozptylem a kovarianční strukturou splňující $\text{cov}(y_t, y_s) = \text{cov}(y_{t+h}, y_{s+h})$ pro libovolné h .

V Boxově–Jenkinsově metodologii můžeme zpracovávat pouze stacionární časové řady. Pokud budeme zpracovávat nestacionární časové řady, pak existují různé způsoby převedení nestacionárních časových řad na stacionární. Tyto metody si uvedeme v podkapitole 2.12.1.

Autokovarianční funkce stacionární časové řady y_t v bodě k je dána vztahem

$$\gamma_k = \text{cov}(y_t, y_{t+k}) = E(y_t - \mu)(y_{t+k} - \mu), \quad k = \dots, -1, 0, 1, \dots,$$

kde $\mu = E(y_t)$. Hodnotu jejího odhadu určíme ze vztahu $\mu = \bar{y} = \frac{1}{n} \sum_{t=1}^n y_t$. Odhad autokovarianční funkce získáme následovně

$$c_k = \frac{1}{n} \sum_{t=1}^{n-k} (y_t - \bar{y})(y_{t+k} - \bar{y}), \quad k = 0, 1, \dots, n-1.$$

Dále budeme potřebovat *autokorelační funkci* ACF, kterou definujeme jako

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \frac{\gamma_k}{\sigma_y^2}, \quad k = \dots, -1, 0, 1, \dots, \quad (11)$$

kde $\sigma_y^2 = \gamma_0 = \text{var}(y_t)$. Odhad ACF získáme užitím autokovariančních odhadů

$$r_k = \frac{c_k}{c_0}, \quad k = 0, 1, \dots, n-1.$$

Do všech uvedených vzorců dosazujeme hodnoty ze stejné časové řady, a proto jsou názvy obecných funkcí doplněny o předponu *auto*. Funkce autokovariance a autokorelace charakterizují sílu závislosti mezi y_t a y_{t+k} bez vlivu ostatních veličin nacházejících se mezi nimi. Obě funkce jsou sudé, a proto se omezíme ve výpočtech pouze na $k \geq 0$. Graficky autokorelační funkci zachycuje korelogram, pro který platí $\rho_0 = 1$ a $|\rho_k| \leq 1$. Abychom při využívání těchto funkcí dosahovali korektních výsledků, je nutné aby časová řada obsahovala alespoň 50 členů a k nepřesahovalo $\frac{n}{4}$.

Chování autokorelační funkce je důležitým ukazatelem při analýze dané časové řady pro odhalení vhodného typu modelu. Při analýze funkce ρ_k hledáme takové k , za kterým začínají hodnoty autokorelační funkce být nulové. Tento bod

k budeme nazývat *identifikační bod* a označíme ho k_0 . U analyzované časové řady neznáme skutečné hodnoty ρ_k , ale pouze jejich odhady r_k . A protože neznáme skutečné hodnoty, musíme hledat mez, kdy budou hodnoty r_k již dostatečně blízké nule, abychom mohli s předem danou spolehlivostí tvrdit, že $\rho_k = 0$. Testujeme hypotézu

$$H_0 : \rho_k = 0, \quad k > k_0.$$

Pro stanovení meze, kterou nesmí hodnoty autokorelační funkce překročit, se užívá Bartlettova aproximace [5]

$$\sigma(r_k) = \sqrt{\text{var}(r_k)} \sim \sqrt{\frac{1}{n} \left(1 + 2 \sum_{j=1}^{k_0} r_j^2 \right)}, \quad k > k_0. \quad (12)$$

Pokud budou absolutní hodnoty odhadů autokorelační funkce menší než dvojnásobek Bartlettovy aproximace, pak aktuální hodnota k je hledaným identifikačním bodem, tzn. jestliže $|r_k| < 2\sigma(r_k)$, pak $k = k_0$. Když některá z hodnot autokorelační funkce překročí tuto mez, pak položíme $k = k_0 + 1$ a znovu otestujeme odhady autokorelační funkce. Takto budeme postupovat do splnění uvedené nerovnosti, nebo do vyčerpání doporučeného počtu k . V případě nepotvrzení nerovnosti prohlásíme, že identifikační bod neexistuje a bude platit $k_0 = \infty$.

Parciální autokorelační funkce ρ_{kk} je parciální korelační koeficient y_t a y_{t+k} při neměnných hodnotách $y_{t+1}, y_{t+2}, \dots, y_{t+k-1}$. Vyjadřuje sílu závislosti mezi y_t a y_{t+k} bez vlivu veličin $y_{t+1}, y_{t+2}, \dots, y_{t+k-1}$. Budeme ji označovat PACF.

Hodnotu parciální autokorelační funkce vypočítáme dosazením do následujícího vzorce [5]

$$\rho_{kk} = \frac{|\mathbf{C}_k|}{|\mathbf{D}_k|}, \quad \text{kde matice autokorelací jsou ve tvaru}$$

$$\mathbf{C}_k = \begin{vmatrix} 1 & \rho_1 & \rho_2 & \dots & \rho_{k-2} & \rho_1 \\ \rho_1 & 1 & \rho_1 & \dots & \rho_{k-3} & \rho_2 \\ \rho_2 & \rho_1 & 1 & \dots & \rho_{k-4} & \rho_3 \\ \rho_3 & \rho_2 & \rho_1 & \dots & \rho_{k-5} & \rho_4 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \dots & \rho_1 & \rho_k \end{vmatrix}, \quad \mathbf{D}_k = \begin{vmatrix} 1 & \rho_1 & \rho_2 & \rho_3 & \dots & \rho_{k-1} \\ \rho_1 & 1 & \rho_1 & \rho_2 & \dots & \rho_{k-2} \\ \rho_2 & \rho_1 & 1 & \rho_1 & \dots & \rho_{k-3} \\ \rho_3 & \rho_2 & \rho_1 & 1 & \dots & \rho_{k-4} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \rho_{k-4} & \dots & 1 \end{vmatrix}.$$

Podíl je na první pohled velice složitý, ale při jeho bližším zkoumání zjistíme, že čitatel zlomku vzniká ze jmenovatele pouhou záměnou posledního sloupce determinantu. Platí $\rho_{11} = \rho_1$. Rekurentně lze vypočítat odhady r_{kk} parciální autokorelační funkce ρ_{kk} podle vzorců [5]

$$r_{11} = r_1,$$

$$r_{kk} = \frac{r_k - \sum_{j=1}^{k-1} r_{k-1j} r_{k-j}}{1 - \sum_{j=1}^{k-1} r_{k-1j} r_j} \quad \text{pro } k > 1,$$

kde $r_{kj} = r_{k-1j} - r_{kk} r_{k-1k-j}$ pro $j = 1, 2, \dots, k-1$.

Také u parciální autokorelační funkce určujeme *identifikační bod* k_0 , který pomáhá odhalit vhodný typ modelu charakterizující nejlépe zkoumanou časovou řadu. Opět hledáme bod k , pro který platí hypotéza

$$H_0 : \quad \rho_{kk} = 0 \quad k > k_0.$$

Zde nám k jeho odhalení poslouží Quenouilleova aproximace [5]

$$\sigma(r_{kk}) \sim \sqrt{\frac{1}{n}}, \quad k > k_0. \quad (13)$$

Budeme testovat platnost nerovnosti $|r_{kk}| < 2\sqrt{\frac{1}{n}}$. Pokud pro testované k platí nerovnost, pak bod k je identifikační bod a platí $k = k_0$. Jestliže je $|r_{kk}| > 2\sqrt{\frac{1}{n}}$, potom položíme $k = k_0 + 1$ a zopakujeme testování. To opakujeme do splnění uvedené nerovnosti, nebo do vyčerpání doporučeného počtu k . Po vyčerpání doporučeného počtu k prohlásíme, že identifikační bod parciální autokorelační funkce neexistuje.

2.6.1. Lineární proces

Lineární proces charakterizuje časová řada tvaru [5]

$$y_t = \varepsilon_t + \psi_1 \varepsilon_{t-1} + \psi_2 \varepsilon_{t-2} + \psi_3 \varepsilon_{t-3} + \dots, \quad (14)$$

kde ψ_j jsou parametry a $\varepsilon_t \sim WN(0, \sigma_\varepsilon^2)$. ε_t je opět bílý šum (WN = white noise), který má tyto vlastnosti $E(\varepsilon_t) = 0$, $var(\varepsilon_t) = \sigma_\varepsilon^2$ a $cov(\varepsilon_t, \varepsilon_s) = 0$, kde $t \neq s$. Jiný způsob zápisu lineárního procesu umožňuje operátor zpětného posunutí B

$$By_t = y_{t-1}.$$

Můžeme jej použít i několikrát, jeho j -té užití zapíšeme

$$B^j y_t = y_{t-j}.$$

Lineární proces zapsaný s využitím operátoru zpětného posunutí je ve tvaru

$$y_t = \psi(B)\varepsilon_t, \quad (15)$$

přičemž

$$\psi(B) = 1 + \sum_{j=1}^{\infty} \psi_j B^j. \quad (16)$$

Podmínka

$$\psi(B) \text{ konverguje pro } |B| \leq 1 \quad (17)$$

je postačující pro existenci lineárního procesu. Díky podmínce (17) je zaručena stacionarita procesu a platí $E(y_t) = 0$, tj. proces je centrováný [5].

Pokud chceme současnou hodnotu y_t vyjádřit pomocí minulých hodnot a současné hodnoty bílého šumu, pak lze za jistých podmínek zapsat lineární proces ve tvaru

$$y_t = \pi_1 y_{t-1} + \pi_2 y_{t-2} + \pi_3 y_{t-3} + \dots + \varepsilon_t. \quad (18)$$

Podmínka, která přepis umožňuje je

$$\pi(B) \text{ konverguje pro } |B| \leq 1. \quad (19)$$

Proces, který umožňuje přepis do tvaru (18), se nazývá *invertibilní*. V invertibilním procesu můžeme použít operátor zpětného posunutí B a dostáváme ekvivalentní zápis procesu (18) ve tvaru

$$\pi(B)y_t = \varepsilon_t, \quad \text{kde } \pi(B) = 1 - \sum_{j=1}^{\infty} \pi_j B^j.$$

Mezi parametry $\psi(B)$ a $\pi(B)$, které představují různé způsoby zápisu lineárního procesu, existuje vztah. Nejdříve ve vztahu (15) dosadíme za ε_t , čímž nám vznikne

$$y_t = \psi(B)\pi(B)y_t, \quad \Rightarrow \quad 1 = \psi(B)\pi(B).$$

Mezi parametry lineárního procesu je tedy vztah inverze, kterou zapíšeme

$$\pi(B) = \psi^{-1}(B).$$

2.7. Proces klouzavých součtů

Proces klouzavých součtů je konečný lineární proces, který zapisujeme ve tvaru [5]

$$y_t = \varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} + \dots + \theta_q\varepsilon_{t-q},$$

kde q vyjadřuje řád procesu, $\varepsilon_t \sim WN(0, \sigma_\varepsilon^2)$ je bílý šum a $\theta_1, \dots, \theta_q$ jsou hledané parametry. Proces klouzavých součtů řádu q označujeme $MA(q)$. Tento proces s využitím symboliky zpětného posunutí B zapisujeme

$$y_t = \theta(B)\varepsilon_t,$$

kde $\theta(B)$ značí operátor klouzavých součtů, který je vyjádřen vzorcem

$$\theta(B) = 1 + \sum_{j=1}^q \theta_j B^j.$$

Podmínku stacionarity (17) proces $MA(q)$ splňuje pro libovolnou volbu jeho parametrů. Lineární proces (14) podmínku (17) splňuje vždy. Proces klouzavých součtů je konečnou verzí lineárního procesu, a proto je proces $MA(q)$ vždy stacionární.

Základními číselnými charakteristikami procesu $MA(q)$ jsou nulová střední hodnota, kterou má díky přítomnosti bílého šumu, a rozptyl σ_y^2 ve tvaru

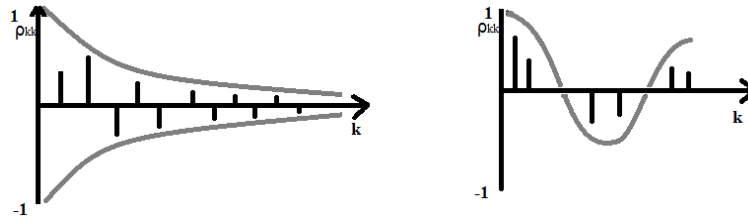
$$\sigma_y^2 = \gamma_0 = (1 + \theta_1^2 + \theta_2^2 + \dots + \theta_q^2)\sigma_\varepsilon^2. \quad (20)$$

Autokorelační funkce procesu klouzavých součtů je dle [5] ve tvaru

$$\rho_k = \begin{cases} \frac{\theta_k + \theta_1\theta_{k+1} + \theta_2\theta_{k+2} + \dots + \theta_{q-k}\theta_q}{1 + \theta_1^2 + \theta_2^2 + \dots + \theta_q^2} & \text{pro } k = 1, \dots, q, \\ 0 & \text{pro } k > q. \end{cases}$$

Výše uvedený zlomek vzniká dosazením do vzorce (11), kde čítec tvoří auto-kovarianční funkce $\gamma_k = \text{cov}(y_t, y_{t+k}) = (\theta_k + \theta_1\theta_{k+1} + \theta_2\theta_{k+2} + \dots + \theta_{q-k}\theta_q)\sigma_\varepsilon^2$ a do jmenovatele dosadíme výraz (20). Identifikační bod autokorelační funkce procesu klouzavých součtů řádu q je $k_0 = q$.

Další charakteristikou procesu $MA(q)$ je *parciální autokorelační funkce*, která v případě procesu $MA(q)$ nemá identifikační bod k_0 . Nemůžeme říct, že by od nějakého bodu byla tato funkce nulová, je pouze omezená geometricky klesající posloupností nebo sinusovkou s geometricky klesající amplitudou. Obě situace jsou zobrazeny na obrázku 1. Vlevo jsou zobrazeny hodnoty parciální autokorelační funkce s geometricky klesající posloupností. V pravé části obrázku 1 je graf parciální autokorelační funkce omezený sinusovkou s geometricky klesající amplitudou.



Obrázek 1: Průběh ρ_{kk} procesu $MA(q)$ (zdroj: [5, str. 114])

Proces $MA(q)$ je *invertibilní*, jestliže umožňuje přepis dle podmínky (18), tedy pokud řada $\pi(B) = \theta^{-1}(B)$ je konvergující pro $|B| \leq 1$. Když označíme M_1, \dots, M_q jako kořeny polynomu $\theta(B)$, potom můžeme polynom $\theta(B)$ rozložit na kořenové činitele [5]

$$\theta(B) = \prod_{j=1}^q (1 - M_j^{-1}B).$$

Řada $\theta^{-1}(B)$ konverguje pro $|B| \leq 1$, pokud platí $|M_j| > 1$, kde $j = 1, \dots, q$.

Neboli když se všechny kořeny polynomu $\theta(B)$ nachází vně jednotkového kruhu, potom proces $MA(q)$ můžeme označit za invertibilní.

Bližší popis procesů $MA(1)$ a $MA(2)$ můžeme najít v [5, str. 110].

2.8. Autoregresní proces

Autoregresní proces řádu p je definován vztahem [5]

$$y_t = \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \dots + \varphi_p y_{t-p} + \varepsilon_t, \quad \varepsilon_t \sim WN(0, \sigma_\varepsilon^2). \quad (21)$$

Proces označujeme zkratkou $AR(p)$. I tento proces můžeme zapsat za využití symboliky operátoru zpětného posunutí B . Zde užíváme autoregresní operátor $\varphi(B)$ ve tvaru

$$\varphi(B) = 1 - \sum_{j=1}^p \varphi_j B^j,$$

pomocí něhož autoregresní proces řádu p přepíšeme do tvaru

$$\varphi(B)y_t = \varepsilon_t.$$

Proces $AR(p)$ je definován v invertibilním tvaru, což si můžeme ověřit při porovnání se vztahem (18). Pokud ho chceme psát ve stacionárním tvaru, pak podmínka stacionarity umožňuje proces $AR(p)$ přepsat do původního neinvertovaného tvaru

$$y_t = \varphi^{-1}(B)\varepsilon_t.$$

Jestliže se všechny kořeny polynomu $\varphi(B)$ nacházejí mimo jednotkový kruh, potom je proces $AR(p)$ stacionární.

Základní charakteristikou stacionárního procesu $AR(p)$ je nulová střední hodnota. Hodnoty *autokorelační funkce* ρ_k procesu $AR(p)$ dle [5] získáme, pokud vztah (21) vynásobíme postupně veličinami y_{t-k} pro $k > 0$ a následně vypočítáme střední hodnotu vztahů. Víme, že střední hodnota procesu $AR(p)$ je nulová, a proto je střední hodnota minulých pozorování a současné chyby také nulová. Můžeme tedy vynechat členy $E(y_{t-k}\varepsilon_t) = 0$ pro $k > 0$. Následně vynásobíme

rovnice výrazem $\frac{1}{\gamma_0}$. Vzniká nám autokorelační funkce ρ_k procesu $AR(p)$, která vyhovuje soustavě diferenčních rovnic

$$\rho_k = \varphi_1 \rho_{k-1} + \varphi_2 \rho_{k-2} + \dots + \varphi_p \rho_{k-p}, \quad k > 0. \quad (22)$$

Na základě teorie diferenčních rovnic dostáváme z (22) tvar autokorelační funkce ρ_k procesu $AR(p)$

$$\rho_k = \alpha_1 R_1^{-k} + \alpha_2 R_2^{-k} + \dots + \alpha_p R_p^{-k}, \quad k \geq 0, \quad (23)$$

kde R_j jsou kořeny polynomu $\varphi(B)$ a α_j jsou konstanty, $j = 1, \dots, p$. Tento tvar autokorelační funkce ρ_k procesu $AR(p)$ užíváme, pokud jsou kořeny R_j reálné a navzájem různé.

V případě, že kořeny R_i, R_l polynomu $\varphi(B)$ tvoří pár komplexně sdružených čísel, potom odpovídající členy nahradíme následovně

$$\alpha_i R_i^{-k} + \alpha_l R_l^{-k} \Rightarrow \alpha d^k \sin(\lambda k + \varphi),$$

kde $d \in (0, 1)$ a $\alpha, d, \lambda, \varphi$ představují konstanty.

Poslední z možností je, že kořeny R_1, \dots, R_p polynomu $\varphi(B)$ nejsou navzájem reálné různé, ale jeden z nich je r násobný kořen, např. R_i . Pak v autokorelační funkci (23) tento r násobný kořen typu $\alpha_i R_i^{-k}$ nahradíme vztahem

$$(\beta_0 + \beta_1 k + \dots + \beta_{r-1} k^{r-1}) R_i^{-k},$$

kde $\beta_0, \dots, \beta_{r-1}$ jsou konstanty. V tomto případě výrazně průběh členu R_i^{-k} překrývá průběh polynomu $\beta_0 + \beta_1 k + \dots + \beta_{r-1} k^{r-1}$. Autokorelační funkce autoregresního procesu řádu p je lineární kombinací geometricky klesajících posloupností a sinusoid různých frekvencí s geometricky klesajícími amplitudami.

Pokud známe hodnoty autokorelační funkce procesu $AR(p)$ a potřebujeme vypočítat parametry $\varphi_1, \dots, \varphi_p$ tohoto procesu, pak se využívá Yuleova-Walkerova

soustava rovnic [5] tvaru

$$\begin{aligned}\rho_1 &= \varphi_1 + \varphi_2\rho_1 + \dots + \varphi_p\rho_{p-1}, \\ \rho_2 &= \varphi_1\rho_1 + \varphi_2 + \dots + \varphi_p\rho_{p-2}, \\ &\vdots \\ \rho_p &= \varphi_1\rho_{p-1} + \varphi_2\rho_{p-2} + \dots + \varphi_p.\end{aligned}$$

Vztahy vycházejí z (22) pro $k = 1, \dots, p$. Řešením soustavy rovnic jsou parametry $\varphi_1, \dots, \varphi_p$ určené pomocí hodnot ρ_1, \dots, ρ_p autokorelační funkce.

Další charakteristikou procesu $AR(p)$ je rozptyl

$$\sigma_y^2 = \gamma_0 = \frac{\sigma_\varepsilon^2}{1 - \varphi_1\rho_1 - \dots - \varphi_p\rho_p}.$$

Identifikační bod *parciální autokorelační funkce* procesu $AR(p)$ je $k_0 = p$, tzn. $\rho_{kk} = 0$ pro $k > p$.

Charakteristiky autoregresních procesů $AR(1)$ a $AR(2)$ nalezneme v literatuře [5, str. 115].

2.9. Smíšené procesy

Model $ARMA(p, q)$ označuje smíšený proces řádu p a q . Vzniká spojením procesu $MA(q)$ a $AR(p)$ a je definován vztahem [5]

$$y_t = \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \dots + \varphi_p y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q}.$$

Jiný způsob zápisu procesu $ARMA(p, q)$ umožňuje symbolika operátoru zpětného posunutí B

$$\varphi(B)y_t = \theta(B)\varepsilon_t, \quad \text{kde } \varepsilon_t \sim WN(0, \sigma_\varepsilon^2). \quad (24)$$

Jak již bylo řečeno $\varphi(B)$ je autoregresní operátor a $\theta(B)$ označuje operátor klouzavých součtů. Podmínky stacionarity a invertibility pro model $ARMA(p, q)$ vychází z teorie lineárních procesů. Smíšený proces má stejnou podmínku invertibility jako proces $MA(q)$ a podmínku stacionarity jako proces $AR(p)$.

Základní vlastností stacionárního procesu $ARMA(p, q)$ je nulová střední hodnota. *Autokorelační funkce* ρ_k procesu $ARMA(p, q)$ vychází z obdobného odvození jako u procesu $AR(p)$. Soustava [5]

$$\rho_k = \varphi_1 \rho_{k-1} + \varphi_2 \rho_{k-2} + \dots + \varphi_p \rho_{k-p}, \quad k > q,$$

vyhovuje stejně jako u procesu $AR(p)$ analogické soustavě diferenčních rovnic. Řešení této soustavy pro $k \geq \max(0, q - p + 1)$ je tvaru (23) pro navzájem různé reálné kořeny autoregresního operátoru. Jestliže je $q \geq p$, potom průběh prvních $q - p + 1$ hodnot $\rho_0, \dots, \rho_{q-p}$ je odlišný od pozdějšího průběhu (23) autokorelační funkce, která je ve tvaru lineární kombinace geometricky klesajících posloupností a sinusoid s geometricky klesající amplitudou.

Parciální autokorelační funkce ρ_{kk} procesu $ARMA(p, q)$ vychází z procesu $MA(q)$ až pro $k \geq \max(1, p - q + 1)$. Zde nelze určit identifikační bod k_0 .

Základní charakteristiku procesu $ARMA(1, 1)$ popisuje literatura [5, str. 119].

2.10. Integrované smíšené modely

Integrované smíšené modely patří v praxi mezi jedny z nejpoužívanějších. Umožňují stochasticky popsat nejen náhodné šoky, ale i trendovou složku. Modelují časové řady se změnami sklonu nebo úrovně, které mohou mít nesystematicky náhodný charakter. Integrované smíšené modely mají podobný postup konstrukce jako modely $ARMA(p, q)$. Rozdílnost spočívá v použití diferencování původní řady a následném určení modelu.

Až dosud jsme požadovali, aby časové řady byly stacionární. Pro integrované smíšené modely nám budou postačovat homogenně nestacionární časové řady. Takové řady nejsou stacionární, ale převedením pomocí diferencování se stacionárními stávají.

Pro integrovaný smíšený model používáme označení $ARIMA(p, d, q)$ [5]. Tento model je definován

$$\varphi(B)w_t = \theta(B)\varepsilon_t, \quad (25)$$

kde

$$w_t = \Delta^d y_t, \quad \varepsilon_t \sim WN(0, \sigma_\varepsilon^2).$$

Parametr d představuje řád diference zkoumaného procesu y_t . Diferencováním časové řady y_t získáváme příslušný stacionární model $ARMA(p, q)$ pro proces w_t .

Pomocí symboliky operátoru zpětného posunutí B můžeme vyjádřit diferenční operátor Δ takto

$$\Delta = 1 - B, \quad \text{kde} \quad \Delta y_t = y_t - y_{t-1} = (1 - B)y_t.$$

Souhrný zápis modelu $ARIMA(p, d, q)$ můžeme provést dvěma způsoby. Buď jej zapíšeme

$$\varphi(B)(1 - B)^d y_t = \theta(B)\varepsilon_t, \quad \text{nebo} \quad \nu(B)y_t = \theta(B)\varepsilon_t,$$

kde operátor

$$\nu(B) = \varphi(B)(1 - B)^d$$

představuje zobecněný autoregresní operátor. Operátor $\nu(B)$ reprezentuje polynom řádu $p + d$ v proměnné B . Pokud je model (25) stacionární, pak má p kořenů vně jednotkového kruhu v komplexní rovině a d jednotkových kořenů.

Konstrukce modelů $ARIMA(p, d, q)$ probíhá tak, že nejdříve ověříme zda daná řada nepotřebuje transformovat, viz kapitola 2.12. Následně danou řadu diferencujeme, poté určíme model $ARMA(p, q)$ včetně počátečních odhadů parametrů. Při diferencování ale ztrácíme právě d prvních hodnot původní řady. Časová řada y_t je dána n hodnotami y_1, \dots, y_n , po diferencování řádu d získáme časovou řadu w_t , která je délky $n - d$ s hodnotami w_{d+1}, \dots, w_n .

Při diferencování časové řady y_t nemá smysl časovou řadu centrovat. Pro integrovaný smíšený model platí [5]

$$\Delta^d(y_t - \bar{y}) = \Delta^d y_t.$$

Modely $ARIMA(p, d, q)$ můžeme trendy modelovat stochasticky, ale také zachytit časové řady s deterministickým trendem. Do vztahu (25) přidáme konstantu δ a vzniká model ve tvaru [5]

$$\varphi(B)w_t = \delta + \theta(B)\varepsilon_t. \tag{26}$$

Vztahem (26) můžeme vyjádřit procesy tvaru

$$\beta_0 + \beta_1 t + \beta_2 t^2 + \dots + \beta_d t^d + y_t,$$

kde $\beta_0 + \beta_1 t + \beta_2 t^2 + \dots + \beta_d t^d$ je polynomický trend řádu d . Časovou řadu y_t vyjadřuje proces $ARIMA(p, d, q)$ tvaru $\varphi(B)\Delta^d y_t = \theta(B)\varepsilon_t$. Model můžeme totiž upravit do tvaru

$$\varphi(B)(d!\beta_d) + \varphi(B)\Delta^d y_t = \delta + \theta(B)\varepsilon_t,$$

kde $d!$ je faktoriál čísla d a položit $\delta = (1 - \varphi_1 - \varphi_2 - \dots - \varphi_p)d!\beta_d$. Proces w_t je proces modelu $ARMA(p, q)$ s nenulovou střední hodnotou.

Výhodnou vlastností modelů $ARIMA(p, d, q)$ je zmírnění požadavku na předpoklad stacionarity časových řad pro užití $ARMA$ modelů.

2.11. Sezónní integrované smíšené modely

Sezónní modelování časových řad užitím Boxovy-Jenkinsovy metodologie je umožněno pomocí sezónních integrovaných smíšených modelů, které značíme $SARIMA$ modely. Teorii této podkapitoly čerpáme z [5]. Sezónnost je užitím modelů $SARIMA$ popisována stochasticky. Máme časovou řadu např. čtvrtletních nebo měsíčních pozorování a řada vykazuje sezónnost. Nejdříve data rozčleníme např. na hodnoty odpovídající pro jaro, léto, podzim a zimu. Následně hledáme vhodný model pro časovou řadu odpovídající hodnotám konkrétní sezónnosti, např. jaro, a dále vytvoříme vhodný model pro každou z těchto sezón zvlášť. Pro modelování pomocí $SARIMA$ modelů potřebujeme tedy dostatečně dlouhou časovou řadu, abychom po rozčlenění analyzované časové řady měli opět přibližně 50 členů v každé z rozčleněných řad.

Pro hodnoty z jedné rozčleněné časové řady tedy vytvoříme model ve tvaru

$$\Phi(B^L)\Delta_L^D y_t = \Theta(B^L)\eta_t, \quad (27)$$

kde D je řád sezónní difference, L je délka periody dané sezóny a η_t je náhodná složka, která není bílým šumem. Časový index t odpovídá hodnotám konkrétní

sezónnosti. Pokud chceme modelem popsat například hodnoty naměřené v jednotlivých letech pro čtvrtletní data, pak je délka periody $L = 4$. Sezónní autoregresní operátor řádu P zapisujeme

$$\Phi(B^L) = 1 - \Phi_1(B^L) - \Phi_2(B^{L \cdot 2}) - \dots - \Phi_P(B^{L \cdot P}).$$

Sezónní operátor klouzavých součtů řádu Q je tvaru

$$\Theta(B^L) = 1 + \Theta_1(B^L) + \Theta_2(B^{L \cdot 2}) + \dots + \Theta_Q(B^{L \cdot Q})$$

a sezónní diferenční operátor zapíšeme

$$\Delta_L = 1 - B^L.$$

Model (27) budeme chápat jako $ARIMA(p, d, q)$ pro jednotlivou sezónní složku a podobné modely vytvoříme i pro ostatní sezóny (např. léto, ...). V jednotlivých modelech se vyskytuje náhodná složka η_t , která by měla být mezi těmito modely korelovaná. Tato složka představuje propojení mezi jednotlivými modely. Předpokládáme, že náhodná složka η_t je také zachycena modelem $ARIMA(p, d, q)$, který zapíšeme

$$\varphi(B)\Delta^d\eta_t = \theta(B)\varepsilon_t. \quad (28)$$

Složka ε_t představuje bílý šum. Spojením modelů (27) a (28) nám vzniká multiplikativní sezónní model

$$\varphi(B)\Phi(B^L)\Delta^d\Delta_L^D y_t = \theta(B)\Theta(B^L)\varepsilon_t$$

řádu $(p, d, q) \times (P, D, Q)_L$, který značíme $SARIMA(p, d, q) \times (P, D, Q)_L$.

Konstrukce těchto modelů se odlišuje od konstrukce modelů $ARIMA(p, d, q)$ pouze v určení sezónního modelu a řádu diferencování d a D . Diferencování opět využíváme k odstranění nestacionarity časové řady, která je vyvolána přítomností trendu a sezónnosti. Postup konstrukce si uvedeme v následující kapitole.

2.12. Výstavba modelů

V této kapitole popíšeme teoretický postup, resp. jednotlivé kroky potřebné ke správnému určení modelu. Při výstavbě modelu musíme postupně projít tři

fáze, kterými jsou identifikace modelu, odhad parametrů modelu a ověření kvality modelu. V této kapitole čerpáme převážně z [2], [5] a [6].

2.12.1. Identifikace modelu

Jde o první fázi celého postupu určení správného modelu. Zde určíme druh a řád procesu. V následujících fázích zpřesníme počáteční odhady parametrů a ověříme správnost určení modelu. Fázi identifikace modelu si ještě rozdělíme na přípravnou fázi a vlastní identifikaci modelu.

V *přípravné fázi* musíme hned během prvního kroku ověřit stacionaritu modelu. Tu můžeme ověřovat

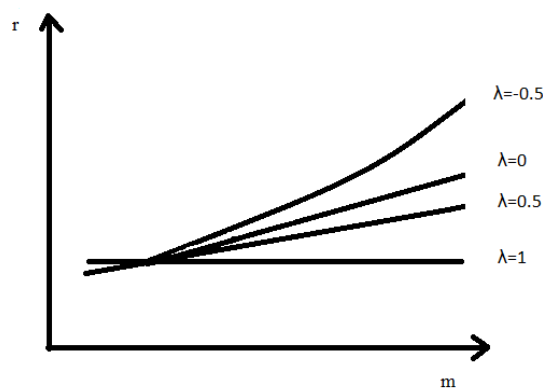
- z grafu funkce, kdy „od oka“ zkoumáme, zda má řada v čase stálý charakter a má konstantní úroveň,
- z definice slabé stacionarity, časovou řadu budeme segmentovat na menší stejně dlouhé části. Vypočítáme z každé této části průměr a rozptyl. Následně hodnoty porovnáváme pomocí párového testu a neparametrických testů.
- z grafů funkcí ACF a PACF - absolutní hodnoty ACF při stacionaritě rychle klesají.

V případě stacionární časové řady se nebudeme zabývat tematikou uvedenou níže, ale přeskočíme až ke zjišťování střední hodnoty stacionárního procesu a ověření, zda je nutné danou časovou řadu centrovat. Pokud zjistíme nestacionaritu zkoumaných dat, pak ji musíme odstranit. Můžeme k tomu využít například diferencování, vyrovnávání dat nebo použít nějaký filtr.

Nejdříve musíme otestovat zkoumanou časovou řadu, zda neobsahuje náhodné šoky, které by neměly charakter bílého šumu s konstantním rozptylem. Test provedeme např. pomocí *mocninné transformace* tvaru [5]

$$y_t^{(\lambda)} = y_t^\lambda \quad \text{pro } \lambda \neq 0,$$
$$y_t^{(\lambda)} = \log y_t \quad \text{pro } \lambda = 0,$$

kde $y_t^{(\lambda)}$ je transformovaná řada. Velikost parametru λ určíme pomocí rozčlenění naší zkoumané časové řady na úseky stejné délky o 4 až 12 pozorováních. V jednotlivých úsecích určíme aritmetický průměr, který označíme m . Vypočítáme si také rozdíl mezi maximální a minimální hodnotou v každém z úseků a označíme ho r . Body se souřadnicemi $[m, r]$ si zaneseme do grafického schématu. Pokud tyto body vyhovují některé z křivek uvedených na obrázku 2, pak budeme dle odpovídajících hodnot řadu y_t transformovat. Pokud leží body v konstantní úrovni ($\lambda = 1$), pak y_t není třeba transformovat. Jestliže body vykazují mírný růst, pak budeme řadu linearizovat odmocňováním. V případě většího růstu budeme používat pro linearizaci y_t logaritmování. Poslední znázorněná křivka ($\lambda = -0.5$) odpovídá linearizování řady pomocí vztahu $\frac{1}{\sqrt{y_t}}$. Transformace se užívají u modelů *ARIMA*(p, d, q) a *SARIMA*. Touto grafickou metodou určíme λ jen přibližně, ale pro naše účely je tento odhad dostačující.



Obrázek 2: Určení transformační konstanty λ (zdroj: [5, str. 145])

Pokud budou hodnoty časové řady stále nestacionární, můžeme použít již zmíněné diferencování pro získání stacionární časové řady. Při využití diferencování během zpracovávání časových řad přecházíme z modelů *ARMA*(p, q) k modelům *ARIMA*. U modelů *ARIMA*(p, d, q) určujeme navíc řád difference d . Existuje hned několik řádů diferencí, a proto si uvedeme vhodné způsoby, jak určit nejvhodnější řád difference. V praxi se nejčastěji používají difference řádu $d = 1$ nebo

$d = 2$. Rozhodnout o vhodném řádu diference můžeme užitím

1. grafů funkce. Graficky si zobrazíme původní nediferencovanou časovou řadu, časovou řadu prvních diferencí a druhých diferencí. Zkoumáme pohledem všechny grafy a rozhodneme o stacionaritě časové řady. Řada musí mít v čase stálý charakter a konstantní úroveň.
2. grafů odhadnuté autokorelační funkce r_k dané řady a její diference. Pokud hodnoty původní autokorelační funkce klesají přibližně lineárně, pak je potřeba diferencovat. Provedeme diferenci a zkoumáme hodnoty autokorelační funkce r_k diferencované řady. Když budou i nadále její hodnoty klesat přibližně lineárně, provedeme další diferenci.
3. odhadnutých rozptylů. V tomto případě budeme porovnávat odhadnutý rozptyl původní řady $\hat{\sigma}_y^2$, rozptyl diferencované řady $\hat{\sigma}_{\Delta y}^2$ a rozptyl druhých diferencí původní řady $\hat{\sigma}_{\Delta^2 y}^2$. Vybereme takovou diferenci d , při které bude nejmenší hodnota rozptylu. Během diferencování se snižují hodnoty odhadovaných rozptylů až do doby, dokud není dosaženo stacionarity. Poté se hodnoty rozptylů opět zvyšují, a proto se doporučuje diferencovat co nejméně.

Po hledání vhodného řádu diferencování mohou nastat dvě situace. Nalezli jsme vhodný řád diference, diferencovaná časová řada se stává stacionární a my přejdeme k vlastní identifikaci modelu. Nebo je vhodný řád diference příliš vysoký, a proto ho nemůžeme použít. Časová řada je tedy stále nestacionární. Pokud má řada sezónní charakter, pak přecházíme k modelům *SARIMA*. V modelech $SARIMA(p, d, q) \times (P, D, Q)_L$ se snažíme odstranit nestacionaritu časové řady způsobenou trendem a sezónním chováním, k čemuž využíváme běžné diferencování a sezónní diferencování. Sezónní diferencování budeme označovat operátorem Δ_L , kde L označuje délku periody v sezónním modelu a D je řád sezónního diferencování. Běžné diferencování značíme operátorem Δ a d určuje řád běžné diference.

Pro správné určení řádu sezónního diferencování zkoumáme odhadnutou autokorelační funkci r_k . Je doporučováno zabývat se čtyřnásobkem délky periody L funkce r_k .

1. Hledáme lokální maxima funkce r_k . Pokud to jsou hodnoty k , které tvoří násobky délky periody L , pak musíme sezónně diferencovat a platí $D = 1$. K sezónnímu diferencování se přistoupilo, protože hodnoty autokorelační funkce v bodech k klesají přibližně lineárně. V případě pomalého poklesu funkce r_k i mezi body k , musíme použít i běžné diferencování.
2. Jinou alternativou určení řádu difference d a D může být nalezení minimální hodnoty z odhadů rozptylů $\hat{\sigma}_y^2, \hat{\sigma}_{\Delta y}^2, \hat{\sigma}_{\Delta_L y}^2, \hat{\sigma}_{\Delta \Delta_L y}^2$. Nejmenší rozptyl určí nejvhodnější řády diferencí.

Jakmile určíme řády diferencí, vzniká nám řada [5]

$$w_t = \Delta^d \Delta_L^D y_t.$$

Nyní pro řadu w_t potřebujeme určit model ve tvaru

$$\varphi(B)\Phi(B^L)w_t = \theta(B)\Theta(B^L)\varepsilon_t.$$

Odhad autokorelační funkce r_k a parciální autokorelační funkce r_{kk} pro řadu w_t nám poslouží ke snadnější identifikaci modelu. Následující dva modely jsou nejuzívanější, ostatní modely *SARIMA* se z důvodu velkého počtu parametrů téměř nevyužívají.

1. Hodnoty r_k jsou významně nenulové v bodech

$$1, \dots, q, L - q, \dots, L + q, 2L - q, \dots, 2L + q, \dots, QL - q, \dots, QL + q,$$

mimo tyto body jsou téměř nulové. Hodnoty parciální autokorelace jsou v úsecích mezi body $L, 2L, \dots$ v absolutní hodnotě geometricky klesající nebo klesají po sinusoidě s geometricky klesající amplitudou. Funkce r_{kk} klesá také v bodech $L, 2L, \dots$. Popsané vlastnosti má pro řadu w_t model $SARIMA(0, 0, q) \times (0, 0, Q)_L$. Při přechodu k řadě y_t dostáváme model $SARIMA(0, d, q) \times (0, D, Q)_L$.

2. Funkce r_k klesá v absolutní hodnotě v úsecích mezi body $L, 2L, \dots$ geometricky nebo po sinusoidě s geometricky klesající amplitudou. Tato funkce klesá také v bodech $L, 2L, \dots$. Funkce r_{kk} je v bodech

$$1, \dots, p, L, \dots, L + p, 2L, \dots, 2L + p, \dots, PL, \dots, PL + p$$

významně nenulová, ale mezi těmito body jsou její hodnoty téměř nulové. Popisu odpovídá model $SARIMA(p, 0, 0) \times (P, 0, 0)_L$ pro řadu w_t . Pro časovou řadu y_t máme model $SARIMA(p, d, 0) \times (P, D, 0)_L$.

Po nalezení vhodného řádu difference jsme získali stacionární časovou řadu. Pokud budeme využívat při zpracování dat diferencování, pak pro nás nebude podstatný krok popsáný v následujícím odstavci, ale přejdeme rovnou k vlastní identifikaci modelu.

Jestliže již původní řada byla stacionární, musíme zjistit střední hodnotu stacionárního procesu. Bude-li tato hodnota nenulová, musíme časovou řadu *centrovat*. Pro ověření nulovosti střední hodnoty, používáme porovnání aritmetického průměru \bar{y} řady y_t a dvojnásobku odhadnuté směrodatné odchylky tohoto průměru $\sigma(\bar{y})$. Pro získání hodnoty $\sigma(\bar{y})$ použijeme následující aproximaci [5]

$$\begin{aligned} \text{var}(\bar{y}) &= \frac{1}{n^2} \text{var} \left(\sum_{t=1}^n y_t \right) = \frac{1}{n^2} \sum_{s=1}^n \sum_{t=1}^n \text{cov}(y_s, y_t) \sim \frac{1}{n} \left(\gamma_0 + 2 \sum_{k=1}^n \gamma_k \right) \\ &\sim \frac{\gamma_0}{n} \left(1 + 2 \sum_{k=1}^{\infty} \rho_k \right). \end{aligned}$$

Pokud je $|\bar{y}| > 2\sigma(\bar{y})$, pak střední hodnota není nulová a musíme řadu *centrovat*. Centrování se provádí odečtením střední hodnoty od jednotlivých hodnot časové řady y_t . V opačném případě můžeme pracovat dále s daty, které necentrujeme. Odhady směrodatných odchylek pro nejčastěji užívané procesy jsou uvedeny v tabulce 1.

Nyní se dostáváme z přípravné fáze k dalšímu kroku identifikace modelu, který nazýváme *vlastní identifikace modelu*. V této části se snažíme zjistit, zda

Tabulka 1: Odhad $\hat{\sigma}(\bar{y})$ (zdroj: [5, str. 123])

model	Odhad $\hat{\sigma}(\bar{y})$
$AR(1)$	$\hat{\sigma}(\bar{y}) = \sqrt{\frac{c_0(1+r_1)}{n(1-r_1)}}$,
$AR(2)$	$\hat{\sigma}(\bar{y}) = \sqrt{\frac{c_0(1+r_1)(1-2r_1^2+r_2)}{n(1-r_1)(1-r_2)}}$,
$MA(1)$	$\hat{\sigma}(\bar{y}) = \sqrt{\frac{c_0(1+2r_1)}{n}}$,
$MA(2)$	$\hat{\sigma}(\bar{y}) = \sqrt{\frac{c_0(1+2r_1+2r_2)}{n}}$,
$ARMA(1, 1)$	$\hat{\sigma}(\bar{y}) = \sqrt{\frac{c_0}{n} \left(1 + \frac{2r_1^2}{r_1-r_2} \right)}$.

existuje identifikační bod k_0 , pomocí kterého bychom určili model a řád zkoumaného procesu. Analyzujeme graf odhadnuté autokorelační a parciální autokorelační funkce. Identifikační bod odhadnuté autokorelační funkce určíme nalezením největšího bodu k , který přísluší pozici hodnoty autokorelační funkce překračující mez danou vzorcem (12). Identifikační bod odhadnuté parciální autokorelační funkce určíme obdobným postupem, ale mez získáme vzorcem (13).

Tabulka 2: Popis tvaru r_k a r_{kk} (zdroj: [3, str. 109])

	$AR(p)$	$MA(q)$	$ARMA(p, q)$
ρ_k	neexistuje k_0 , ρ_k má exponenciálně klesající tvar nebo jde o sinusovku s exponenciálně klesající amplitudou	$k_0 = q$ po q posunutích výrazně klesá	neexistuje k_0 , ρ_k po $q - p$ hodnotách exponenciálně klesající nebo sinusovka s exponenciálně klesající amplitudou
ρ_{kk}	$k_0 = p$ po p posunutích výrazně klesá	neexistuje k_0 , ρ_{kk} má exponenciálně klesající tvar nebo jde o sinusovku s exponenciálně klesající amplitudou	neexistuje k_0 , ρ_{kk} po $p - q$ hodnotách exponenciálně klesající nebo sinusovka s exponenciálně klesající amplitudou

Po určení identifikačních bodů využijeme údajů z tabulky 2 a pokusíme se identifikovat vhodný model. Pracujeme stále pouze s odhady, proto je vhodnější vyzkoušet několik různých alternativ pro správné určení modelu.

U zvoleného modelu musíme určit počáteční odhady parametrů, k tomu využijeme vzorců z tabulky 3. Do vztahů dosadíme odhadnuté hodnoty autokorelační funkce. Získáváme počáteční odhady parametrů vybraného modelu. Jde však pouze o počáteční a nepřesné odhady, protože jsme k jejich určení použili pouze odhady r_k . Jako startovací body iteračních procedur jsou ale postačující, protože je budeme postupně zpřesňovat. V tabulce 3 jsou uvedena také omezení, která by nám mohla pomoci určit správný model. Omezení však nemusí vždy platit, protože stále pracujeme s odhady skutečných hodnot.

Tabulka 3: Počáteční odhady parametrů a omezení na r_k (zdroj: [5, str. 124])

Proces	Počáteční odhady parametrů	Omezení na r_k
$AR(1)$	$\hat{\varphi}_1 = r_1, \quad \hat{\sigma}_\varepsilon^2 = c_0(1 - \hat{\varphi}_1 r_1)$	$ r_1 < 1$
$AR(2)$	$\hat{\varphi}_1 = \frac{r_1(1-r_2)}{1-r_1^2}, \quad \hat{\varphi}_2 = \frac{r_2-r_1^2}{1-r_1^2},$ $\hat{\sigma}_\varepsilon^2 = c_0(1 - \hat{\varphi}_1 r_1 - \hat{\varphi}_2 r_2)$	$ r_2 < 1, \quad r_1^2 < \frac{r_2+1}{2}$
$MA(1)$	$\hat{\theta}_1 = \frac{1-\sqrt{1-4r_1^2}}{2r_1}, \quad \hat{\sigma}_\varepsilon^2 = \frac{c_0}{1+\hat{\theta}_1^2}$	$ r_1 < \frac{1}{2}$
$MA(2)$	$\hat{\theta}_1 = \hat{\theta}_2 = 0.1, \quad \hat{\sigma}_\varepsilon^2 = \frac{c_0}{1+\hat{\theta}_1^2+\hat{\theta}_2^2}$	$r_2 + r_1 > -\frac{1}{2}, \quad r_2 - r_1 > -\frac{1}{2},$ $r_1^2 < 4r_2(1 - 2r_2)$
$ARMA(1, 1)$	$\hat{\varphi}_1 = \frac{r_2}{r_1}, \quad \hat{\theta}_1 = \frac{\hat{b} \pm \sqrt{\hat{b}^2 - 4}}{2},$ $\hat{\sigma}_\varepsilon^2 = \frac{c'_0}{1+\hat{\theta}_1^2},$ kde $\hat{b} = \frac{1-2r_2+\hat{\varphi}_1^2}{r_1-\hat{\varphi}_1}, \quad \hat{\theta}_1 < 1,$ $c'_0 = \sum_{t=1}^n \frac{(y'_t - \bar{y}')^2}{n},$ $y'_t = y_t - \hat{\varphi}_1 y_{t-1}, \quad \bar{y}' = \sum_{t=1}^n \frac{y'_t}{n}$	$2r_1^2 - r_1 < r_2 < r_1 $

2.12.2. Odhad parametrů modelu

V této podkapitole se budeme zabývat iteračním zpřesňováním hrubých odhadů parametrů. Můžeme k tomu využít několik různých postupů, které jsou však aplikovatelné jen s využitím počítačových programů. Představíme si alespoň základní myšlenku těchto metod.

Metoda, která je nejvyužívanější a stála u zrodu několika jejích aplikací, se jmenuje *metoda nejmenších nelineárních čtverců*. Postup si předvedeme na modelu $ARMA(p, q)$, který je dán vztahem (24). Odhady parametrů

$$(\boldsymbol{\varphi}, \boldsymbol{\theta}) = (\varphi_1, \dots, \varphi_p, \theta_1, \dots, \theta_q)$$

modelu $ARMA(p, q)$ určíme minimalizováním součtu čtverců [5]

$$S(\boldsymbol{\varphi}, \boldsymbol{\theta}) = \sum_{t=1}^n \varepsilon_t^2(\boldsymbol{\varphi}, \boldsymbol{\theta}), \quad \text{kde} \quad \varepsilon_t(\boldsymbol{\varphi}, \boldsymbol{\theta}) = \theta^{-1}(B)\varphi(B)y_t.$$

Výraz $\varepsilon_t(\boldsymbol{\varphi}, \boldsymbol{\theta})$ nazýváme odhadnutou hodnotou bílého šumu při daných hodnotách parametrů $\boldsymbol{\varphi}$ a $\boldsymbol{\theta}$. Pro minimalizaci $S(\boldsymbol{\varphi}, \boldsymbol{\theta})$ používáme takový obor parametrů $\boldsymbol{\varphi}$ a $\boldsymbol{\theta}$, aby daný model byl stacionární a invertibilní.

Vycházíme z hodnot časové řady y_t , kde $t = 1, \dots, n$ a odhadujeme parametry $\boldsymbol{\varphi}$ a $\boldsymbol{\theta}$ modelu $ARMA(p, q)$. Pokud model nemá nulovou střední hodnotu, pak budeme hodnoty y_1, \dots, y_n centrovat a odhadovat také parametr μ . V případě, že máme centrovanou časovou řadu, použijeme pro výpočet $\varepsilon_t(\boldsymbol{\varphi}, \boldsymbol{\theta})$ rekurentní vzorec [5]

$$\varepsilon_t(\boldsymbol{\varphi}, \boldsymbol{\theta}) = y_t - \varphi_1 y_{t-1} - \dots - \varphi_p y_{t-p} - \theta_1 \varepsilon_{t-1}(\boldsymbol{\varphi}, \boldsymbol{\theta}) - \dots - \theta_q \varepsilon_{t-q}(\boldsymbol{\varphi}, \boldsymbol{\theta}).$$

Nyní jsme se dostali do bodu, kdy potřebujeme zahájit rekurentní výpočty v čase $t = 1$. Potřebujeme k tomu ovšem určit nějaké startovací hodnoty, tj.

$$y_{1-p}, y_{2-p}, \dots, y_0 \quad \text{a} \quad \varepsilon_{1-q}(\boldsymbol{\varphi}, \boldsymbol{\theta}), \varepsilon_{2-q}(\boldsymbol{\varphi}, \boldsymbol{\theta}), \dots, \varepsilon_0(\boldsymbol{\varphi}, \boldsymbol{\theta}).$$

Tyto hodnoty v současnosti neznáme, ale můžeme je určit pomocí *podmíněné metody nejmenších nelineárních čtverců*, nebo *nepodmíněné metody nejmenších nelineárních čtverců*.

- *Podmíněnou metodu nejmenších nelineárních čtverců* [5] užíváme ve dvou alternativách. Jako první si uvedeme metodu, kdy všechny neznámé hodnoty jsou nulové, tj.

$$y_{1-p} = \dots = y_0 = 0 \quad i \quad \varepsilon_{1-q}(\varphi, \theta) = \dots = \varepsilon_0(\varphi, \theta) = 0.$$

Pokud se kořeny $\varphi(B)$ nachází blízko jednotkové kružnice, pak se mezi hodnotami $y_0 = 0$, kterou jsme si zadali, a reálnou y_1 vznikne velký skok. Ten neodpovídá definovanému procesu a zkreslí celou proceduru odhadování. Často se tedy přistupuje k následující modifikaci. Počítáme hodnoty

$$\varepsilon_{p+1}(\varphi, \theta), \dots, \varepsilon_n(\varphi, \theta)$$

a ostatní hodnoty položíme rovny nule

$$\varepsilon_{p-q+1}(\varphi, \theta) = \dots = \varepsilon_p(\varphi, \theta) = 0.$$

V součtu $S(\varphi, \theta)$ nám sice chybí prvních p sčítanců, ale při rekurentních výpočtech vycházíme ze skutečně naměřených hodnot y_1, \dots, y_p . Podmíněnost této metody spočívá právě v závislosti na počátečních hodnotách řady a bílého šumu.

- *Nepodmíněná metoda nejmenších nelineárních čtverců* [5] eliminuje závislost na počátečních hodnotách. To je příčinou dosahování kvalitnějších odhadů parametrů než u předchozí metody. Minimalizujeme součet čtverců

$$S(\varphi, \theta) = \sum_{t=-\infty}^n [\varepsilon_t(\varphi, \theta)]^2,$$

kde výraz $[\varepsilon_t(\varphi, \theta)]$ představuje podmíněnou střední hodnotu veličiny ε_t při pevných hodnotách y_1, \dots, y_n , a φ i θ jsou hodnoty parametrů uvažovaného modelu. Postup konstrukce pomocí této metody je uveden v [5, str. 130].

Zda použijeme podmíněnou nebo nepodmíněnou metodu nejmenších nelineárních čtverců posuzujeme dle délky časové řady a dle typu určeného modelu. Pro modely s dostatečným počtem pozorování (alespoň 75 pozorování) použijeme podmíněný přístup. V případě, že sledovaná data mají sezónní charakter a aplikujeme na ně model *SARIMA*, použijeme obtížnější nepodmíněný přístup.

Minimalizace součtu nelineárních čtverců je metoda, kterou pro modely s nízkým počtem parametrů získáme jako výstup grafický průběh funkce $S(\boldsymbol{\varphi}, \boldsymbol{\theta})$ a nalezneme globální minimum této funkce. Minimalizaci výrazu $S(\boldsymbol{\varphi}, \boldsymbol{\theta})$ získanou podmíněnou nebo nepodmíněnou metodou přes přípustný obor parametrů dostaneme užitím vhodného počítačového programu. Program vytvoří odpovídající mříž hodnot parametrů $\boldsymbol{\varphi}$ a $\boldsymbol{\theta}$ a křivkami spojí takové uzly mříže, pro které nabývá funkce $S(\boldsymbol{\varphi}, \boldsymbol{\theta})$ přibližně stejných hodnot.

Minimalizovat funkci $S(\boldsymbol{\varphi}, \boldsymbol{\theta})$ budeme aplikováním Gaussova–Newtonova iteračního algoritmu pro podmíněný součet čtverců [5]

$$S(\boldsymbol{\varphi}, \boldsymbol{\theta}) = \sum_{t=p+1}^n \varepsilon_t^2(\boldsymbol{\varphi}, \boldsymbol{\theta}).$$

Počátečním odhadem vektorového parametru $\boldsymbol{\beta} = (\varphi_1, \dots, \varphi_p, \theta_1, \dots, \theta_q)'$ budeme označovat $\boldsymbol{\beta}^0$, který již známe z první fáze výstavby modelů. Jde o startovací bod při iteračním určování modelu. Aplikací Taylorova rozvoje funkcí $\varepsilon_t(\boldsymbol{\varphi}, \boldsymbol{\theta}) = \varepsilon_t(\boldsymbol{\beta})$ v bodě $\boldsymbol{\beta}^0$ ukončeném v lineárním členu získáváme vztah [5]

$$\varepsilon_t(\boldsymbol{\beta}) = \varepsilon_t(\boldsymbol{\beta}^0) - \sum_{i=1}^{p+q} (\beta_i - \beta_i^0) x_{ti}(\boldsymbol{\beta}^0) \quad \text{pro } t = p+1, \dots, n,$$

kde

$$x_{ti}(\boldsymbol{\beta}^0) = -\frac{\partial \varepsilon_t(\boldsymbol{\beta}^0)}{\partial \beta_i}.$$

Přepisem modelu do maticového tvaru lineárního regresního modelu nám vzniká zápis

$$\begin{pmatrix} \varepsilon_{p+1}(\boldsymbol{\beta}^0) \\ \vdots \\ \varepsilon_n(\boldsymbol{\beta}^0) \end{pmatrix}_{(n-p) \times 1} = \begin{pmatrix} -\frac{\partial \varepsilon_{p+1}(\boldsymbol{\beta}^0)}{\partial \beta_1} & \cdots & -\frac{\partial \varepsilon_{p+1}(\boldsymbol{\beta}^0)}{\partial \beta_{p+q}} \\ \vdots & \ddots & \vdots \\ -\frac{\partial \varepsilon_n(\boldsymbol{\beta}^0)}{\partial \beta_1} & \cdots & -\frac{\partial \varepsilon_n(\boldsymbol{\beta}^0)}{\partial \beta_{p+q}} \end{pmatrix}_{(n-p) \times (p+q)} \begin{pmatrix} \boldsymbol{\beta} - \boldsymbol{\beta}^0 \end{pmatrix}_{(p+q) \times 1} + \begin{pmatrix} \varepsilon_{p+1}(\boldsymbol{\beta}) \\ \vdots \\ \varepsilon_n(\boldsymbol{\beta}) \end{pmatrix}_{(n-p) \times 1}.$$

Lineární regresní model je tedy tvaru

$$\varepsilon(\beta^0) = \mathbf{X}(\beta^0)(\beta - \beta^0) + \varepsilon(\beta).$$

Regresí $\varepsilon(\beta^0)$ na $\mathbf{X}(\beta^0)$ dostáváme odhad optimálního přírůstku parametrů $\widehat{\beta - \beta^0}$, tj. [5]

$$\widehat{\beta - \beta^0} = (\mathbf{X}'(\beta^0)\mathbf{X}(\beta^0))^{-1}\mathbf{X}'(\beta^0)\varepsilon(\beta^0).$$

Poté postupně iteračně zpřesňujeme odhady β dosazováním do výrazu

$$\beta^1 = \beta^0 + \widehat{\beta - \beta^0}.$$

Odhad $\widehat{\beta - \beta^0}$ můžeme modifikovat použitím jednotkové matice \mathbf{I} a parametru ν . Následující vzorec nám umožní zrychlit konvergenci Gaussova–Newtonova iteračního algoritmu a zlepšit podmíněnost matice $\mathbf{X}'(\beta^0)\mathbf{X}(\beta^0)$

$$\widehat{\beta - \beta^0} = (\mathbf{X}'(\beta^0)\mathbf{X}(\beta^0) + \nu\mathbf{I})^{-1}\mathbf{X}'(\beta^0)\varepsilon(\beta^0),$$

kde neznámý parametr ν se v průběhu iterace vhodně mění.

Pomocí Gaussovy–Newtonovy metody lze odhadnout rozptylovou matici výsledného odhadu $\hat{\beta}$ ve tvaru

$$\text{var}(\hat{\beta}) = \hat{\sigma}_\varepsilon^2(\mathbf{X}'(\hat{\beta})\mathbf{X}(\hat{\beta}))^{-1}. \quad (29)$$

Odhadnutý rozptyl bílého šumu je ve tvaru $\hat{\sigma}_\varepsilon^2 = \frac{S(\hat{\beta})}{n}$, nebo $\hat{\sigma}_\varepsilon^2 = \frac{S(\hat{\beta})}{n-(p+q)}$. S využitím vztahu (29) získáme aproximace odhadů směrodatných odchylek odhadnutých parametrů u nejvyžívanějších modelů, které shrnuje tabulka 4.

2.12.3. Ověřování modelu

Poslední fází výstavby modelu je ověřování kvality zvoleného modelu v předchozích fázích. Tato fáze potvrdí vhodnost zvoleného modelu, nebo ho zamítne jako neadekvátní a určí nám, jak postupovat při jeho opravě. Poskytne nám tedy informace, jaký model bude vhodnější. Uvedeme si několik metod pro ověření modelu.

Tabulka 4: Aproximace σ odhadnutých parametrů (zdroj: [5, str. 134])

Proces	Směrodatné odchylky odhadnutých parametrů
$AR(1)$	$\sigma(\hat{\varphi}_1) \sim \sqrt{\frac{1-\hat{\varphi}_1^2}{n}}$
$AR(2)$	$\sigma(\hat{\varphi}_1) \sim \sqrt{\frac{1-\hat{\varphi}_2^2}{n}} \sim \sigma(\hat{\varphi}_2)$
$MA(1)$	$\sigma(\hat{\theta}_1) \sim \sqrt{\frac{1-\hat{\theta}_1^2}{n}}$
$MA(2)$	$\sigma(\hat{\theta}_1) \sim \sqrt{\frac{1-\hat{\theta}_2^2}{n}} \sim \sigma(\hat{\theta}_2)$
$ARMA(1, 1)$	$\sigma(\hat{\varphi}_1) \sim \sqrt{\frac{(1-\hat{\varphi}_1^2)(1+\hat{\varphi}_1\hat{\theta}_1)^2}{n(\hat{\varphi}_1+\hat{\theta}_1)^2}}, \quad \sigma(\hat{\theta}_1) \sim \sqrt{\frac{(1-\hat{\theta}_1^2)(1+\hat{\varphi}_1\hat{\theta}_1)^2}{n(\hat{\varphi}_1+\hat{\theta}_1)^2}}$

Každá metoda zhodnotí daný model s různou účinností a z jiného hlediska, proto se doporučuje aplikovat na získaný model více metod.

Metoda přeparametrizování modelu je první z uvedených metod. Pokud si nejsme jisti kvalitou zvoleného modelu a zdají se nám velké směrodatné odchylky odhadnutých parametrů, nebo vysoké hodnoty $\hat{\sigma}_\varepsilon^2$, pak můžeme přidat další parametr. Zkusíme použít model s více parametry a jeho vhodnost ověříme tak, že se musí zmenšit směrodatné odchylky odhadnutých parametrů i rozptylu bílého šumu $\hat{\sigma}_\varepsilon^2$. Nové parametry musí být odlišné od nuly.

Metoda odhadnutých reziduí [5] pracuje s odhadnutými hodnotami bílého šumu určenými ve fázi odhadování parametrů

$$\hat{\varepsilon}_t = \varepsilon_t(\hat{\varphi}, \hat{\theta}).$$

Metoda se zabývá individuálním zkoumáním korelačního chování odhadnutých reziduí $\hat{\varepsilon}_t$. Pro model $ARMA(p, q)$ platí

$$\hat{\varepsilon}_t = \hat{\theta}^{-1}(B)\hat{\varphi}(B)y_t = y_t - \hat{\varphi}_1 y_{t-1} - \dots - \hat{\varphi}_p y_{t-p} - \hat{\theta}_1 \hat{\varepsilon}_{t-1} - \dots - \hat{\theta}_q \hat{\varepsilon}_{t-q}.$$

Odhadnuté polynomy $\hat{\varphi}(B)$ a $\hat{\theta}(B)$ získáme dosazením odhadnutých parametrů do $\varphi(B)$ a $\theta(B)$. Z odhadnutých reziduí $\hat{\varepsilon}_t$ vznikla časová řada, pro kterou můžeme vypočítat odhadnuté hodnoty autokorelační funkce $r_k(\hat{\varepsilon})$. Pro potřeby dalšího popisu si ještě označíme odhadnuté hodnoty autokorelační funkce bílého šumu ε_t jako $r_k(\varepsilon)$.

Pokud náhodné veličiny ε_t mají normální rozdělení $N(0, 1)$ a jsou nezávislé, pak odhadnuté hodnoty $r_k(\varepsilon)$ pro větší n můžeme aproximovat nezávislými náhodnými veličinami, které mají rozdělení $N(0, \frac{1}{n})$.

Při dosazení odhadnutých hodnot autokorelační funkce $r_k(\hat{\varepsilon})$ za neznámé hodnoty autokorelační funkce bílého šumu $r_k(\varepsilon)$ není zachována pro malé hodnoty k nezávislost a rozptyl tvaru $\frac{1}{n}$. Malé hodnoty k narozdíl od velkých hodnot n zachovávají pouze normalitu a nulovou střední hodnotu. Pro zachování rozptylu byl odvozen pro model $AR(1)$ vztah

$$\text{var}\{r_k(\hat{\varepsilon})\} \sim \frac{1 - \varphi_1^{2(k-1)}(1 - \varphi_1^2)}{n},$$

který se pro větší k přiblíží k hodnotě $\frac{1}{n}$. Pro hodnotu $r_1(\hat{\varepsilon})$ platí vztah

$$\text{var}\{r_1(\hat{\varepsilon})\} \sim \frac{\varphi_1^2}{n}.$$

Hodnota $\frac{\varphi_1^2}{n}$ může být značně menší než hodnota $\frac{1}{n}$. Pro $AR(2)$ platí vztahy

$$\text{var}\{r_1(\hat{\varepsilon})\} \sim \frac{\varphi_2^2}{n} \quad \text{a} \quad \text{var}\{r_2(\hat{\varepsilon})\} \sim \frac{\varphi_2^2 + \varphi_1^2(1 + \varphi_2)^2}{n}.$$

Pro procesy $MA(q)$ a $ARMA(p, q)$ můžeme odvodit obdobné vztahy jako pro proces $AR(p)$. Při určování vlastností $r_k(\hat{\varepsilon})$ u procesu $ARMA(p, q)$ definovaného vztahem $\varphi(B)y_t = \theta(B)\varepsilon_t$ můžeme proces chápat jako proces $AR(p+q)$ tvaru $\varphi(B)\theta(B)y_t = \varepsilon_t$. Důvodem je, že odhadnutá rezidua jsou v procesech $AR(1)$ a $AR(2)$ přibližně stejná. Pro proces $MA(1)$ je rozptyl $r_k(\hat{\varepsilon})$ tvaru

$$\text{var}\{r_k(\hat{\varepsilon})\} \sim \frac{1 - \theta_1^{2(k-1)}(1 - \theta_1^2)}{n}.$$

Metoda odhadnutých reziduí srovnává hodnoty $r_k(\hat{\varepsilon})$ s dvojnásobkem jejich směrodatných odchylek, které získáme ze vztahu

$$\sigma\{r_k(\hat{\varepsilon})\} = \sqrt{\text{var}\{r_k(\hat{\varepsilon})\}}.$$

Model můžeme označit za nevhodně určený, pokud některá hodnota $r_k(\hat{\varepsilon})$ bude v absolutní hodnotě větší než dvojnásobek směrodatné odchylky $\sigma\{r_k(\hat{\varepsilon})\}$.

Více používaným testem než předchozí Metoda odhadnutých reziduí je *Portmanteau test*. Tento test posuzuje vhodnost zvoleného modelu užitím souhrného korelačního chování odhadnutých reziduí $\hat{\varepsilon}_t$. Dává sice méně přesné výsledky než předchozí metoda, ale podstatně se snížila jeho výpočetní složitost. Pro Portmanteau test musíme vypočítat tzv. portmanteau statistiku [5]

$$Q = n \sum_{k=1}^K r_k^2(\hat{\varepsilon}),$$

kde n je délka zkoumané řady, hodnota K je vhodným způsobem vybrané číslo. Číslo K se hojně zaměňuje s hodnotou \sqrt{n} . Pokud budeme předpokládat, že pomocí Portmanteau testu ověřujeme adekvátnost modelu $ARMA(p, q)$, pak portmanteau statistika Q má pro řady s větším počtem členů přibližně rozdělení χ_{K-p-q}^2 . Na dané hladině významnosti α zamítneme model $ARMA(p, q)$ jako neadekvátní v případě, že statistika Q překročí příslušnou kritickou hodnotu $\chi_{K-p-q}^2(\alpha)$.

Modifikací statistiky Q je Ljungova–Boxova statistika [5]

$$Q^* = n(n+2) \sum_{k=1}^K \frac{r_k^2(\hat{\varepsilon})}{n-k}, \quad (30)$$

která dává kvalitnější výsledky než statistika Q . Statistiku Q^* opět porovnáваме na hladině významnosti α s kritickou hodnotou $\chi_{K-p-q}^2(\alpha)$. Pro zvolený model budou svědčit hodnoty nižší než daná kritická hodnota.

Pokud jsme určili více vhodných modelů a potřebujeme rozhodnout, který z nich je nejlepší, použijeme následující kritéria.

Akaikeho informační kritérium AIC pomáhá rozhodnout, který z určených modelů je vhodnější. Vybereme model, který má nižší hodnotu *AIC*. Statistika [6]

$$AIC(M) = \ln \hat{\sigma}_m^2 + \frac{2M}{n} \quad (31)$$

zhodnotí určené modely, kde M je součet parametrů v modelu $ARMA(p, q)$, $\hat{\sigma}_m^2$ představuje reziduální rozptyl daného modelu, n je počet pozorování.

Při výběru vhodného modelu je užitečné porovnávat výsledky více kritérií. Jako jedno z dalších možných hledisek našeho výběru si uvedeme úpravu předchozího kritéria – *Modifikované Akaikeho kritérium AICC*, které je tvaru [2]

$$AICC(M) = \ln \hat{\sigma}_m^2 + \frac{\frac{2M}{1 - \frac{M+1}{n}}}{n},$$

kde $M = p + q$ je z modelu $ARMA(p, q)$, $\hat{\sigma}_m^2$ charakterizuje reziduální rozptyl určeného modelu. Pro rostoucí k první člen součtu klesá a druhý, který penalizuje větší počet regresorů, roste. Model, který vybereme jako vhodnější, bude ten s nejnižší hodnotou *AICC*.

Posledním kritériem, které si uvedeme, je *Schwarzovo informační kritérium* označované *BIC* nebo také *SC*. Je jinou alternativou pro porovnání modelů [6]

$$SC(M) = \ln \hat{\sigma}_m^2 + \frac{M \ln n}{n}.$$

Za nejlepší označíme opět model s nejnižší hodnotou *SC*.

2.13. Předpovědi

Po správné identifikaci modelu a ověření adekvátnosti modelu již není obtížné zkonstruovat predikce. Postup konstrukce si ukážeme na modelu $ARMA(p, q)$, který je ve tvaru

$$y_t = \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \dots + \varphi_p y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q}.$$

Předpověď hodnoty y_{t+h} v čase t budeme značit $\hat{y}_{t+h}(t)$ a interpretovat jako předpověď v čase t o h kroků dopředu. Budeme konstruovat $\hat{y}_{t+h}(t)$ jako lineární předpověď, která je lineární funkcí hodnot y_t, y_{t-1}, \dots nebo hodnot $\varepsilon_t, \varepsilon_{t-1}, \dots$. Pro předpověď musí platit, že má ve třídě všech lineárních předpovědí nejmenší střední čtvercovou chybou $E(y_{t+h} - \hat{y}_{t+h}(t))^2$. Předpověď dané časové řady budeme hledat ve tvaru [5]

$$\hat{y}_{t+h}(t) = \psi_h^* \varepsilon_t + \psi_{h+1}^* \varepsilon_{t-1} + \dots,$$

podle symboliky užívané v modelech $ARMA(p, q)$ můžeme psát

$$y_{t+h} = \varepsilon_{t+h} + \psi_1 \varepsilon_{t+h-1} + \psi_2 \varepsilon_{t+h-2} + \dots + \psi_{h-1} \varepsilon_{t+1} + \psi_h \varepsilon_t + \dots$$

Snažíme se tedy určit koeficienty $\psi_h^*, \psi_{h+1}^*, \psi_{h+2}^*, \dots$, které po dosazení do výrazu

$$\left(1 + \psi_1^2 + \dots + \psi_{h-1}^2 + \sum_{j=h}^{\infty} (\psi_j - \psi_j^*)^2\right) \sigma_\varepsilon^2$$

hodnotu tohoto výrazu minimalizují. Výraz má minimální hodnotu, pokud platí $\psi_j^* = \psi_j$, kde $j = h, h+1, \dots$. Z uvedeného vyplývá, že předpověď je ve tvaru $\hat{y}_{t+h}(t) = \psi_h \varepsilon_t + \psi_{h+1} \varepsilon_{t-1} + \dots$ a můžeme ji tedy zapsat jako [5]

$$\hat{y}_{t+h}(t) = [y_{t+h}], \quad (32)$$

kde výraz v hranatých závorkách na pravé straně rovnice označuje podmíněnou střední hodnotu při pevných hodnotách y_t, y_{t-1}, \dots

Při určování budoucích hodnot vždy vznikají chyby. Chybu předpovědi $\hat{y}_{t+h}(t)$ definujeme jako [5]

$$\epsilon_{t+h}(t) = y_{t+h} - \hat{y}_{t+h}(t).$$

Po dosazení za y_{t+h} a $\hat{y}_{t+h}(t)$ získáváme chybu předpovědi ve tvaru

$$\epsilon_{t+h}(t) = \varepsilon_{t+h} + \psi_1 \varepsilon_{t+h-1} + \dots + \psi_{h-1} \varepsilon_{t+1}.$$

Chyby předpovědi mají nulovou střední hodnotu, tj. $E(\epsilon_{t+h}(t)) = 0$, a rozptyl ve tvaru

$$\text{var}(\epsilon_{t+h}(t)) = (1 + \psi_1^2 + \dots + \psi_{h-1}^2) \sigma_\varepsilon^2. \quad (33)$$

Speciálně platí [5]

$$\epsilon_t(t-1) = y_t - \hat{y}_t(t-1) = \varepsilon_t.$$

2.13.1. Konstrukce předpovědí

V této podkapitole čerpáme z [5, str. 155]. Při konstrukci předpovědí vycházíme z modelu

$$y_{t+h} = \varphi_1 y_{t+h-1} + \dots + \varphi_p y_{t+h-p} + \varepsilon_{t+h} + \theta_1 \varepsilon_{t+h-1} + \dots + \theta_q \varepsilon_{t+h-q},$$

který díky následujícím vztahům

$$\begin{aligned} [y_{t+j}] &= \hat{y}_{t+j}(t) && \text{pro } j > 0, \\ [y_{t+j}] &= y_{t+j} && \text{pro } j \leq 0, \\ [\varepsilon_{t+j}] &= 0 && \text{pro } j > 0, \\ [\varepsilon_{t+j}] &= \varepsilon_{t+j} = y_{t+j} - \hat{y}_{t+j}(t+j-1) && \text{pro } j \leq 0, \end{aligned}$$

můžeme přepsat do tvaru

$$\hat{y}_{t+h}(t) = \varphi_1 [y_{t+h-1}] + \dots + \varphi_p [y_{t+h-p}] + [\varepsilon_{t+h}] + \theta_1 [\varepsilon_{t+h-1}] + \dots + \theta_q [\varepsilon_{t+h-q}]. \quad (34)$$

Z tohoto vztahu vycházíme při výpočtech předpovědí.

Postup tvorby předpovědí:

1. Při výpočtech předpovědí postupujeme rekurentně, nejdříve vypočítáme předpovědi o jeden krok dopředu $\hat{y}_{q+1}(q), \hat{y}_{q+2}(q+1), \dots$ a z nich předpovědi o další krok dopředu a takto postupujeme pořád dál.
2. Za hodnoty na pravé straně vzorce (34) dosazujeme následovně

$$\begin{aligned} [\varepsilon_s] &= 0 && \text{pro } s > t, \\ [y_s] &= \hat{y}_s(t) && \text{pro } s > t, \\ [\varepsilon_s] &= y_s - \hat{y}_s(s-1) && \text{pro } s \leq t, \\ [y_s] &= y_s && \text{pro } s \leq t. \end{aligned}$$

3. Pro rekurentní výpočty musíme položit $\varepsilon_1 = \varepsilon_2 = \dots = \varepsilon_q = 0$.

95% předpovědní interval v čase t o h kroků dopředu pro proces $ARMA(p, q)$ je interval

$$\left(\hat{y}_{t+h}(t) - 2\sigma\{\varepsilon_{t+h}(t)\}; \hat{y}_{t+h}(t) + 2\sigma\{\varepsilon_{t+h}(t)\} \right), \quad (35)$$

do kterého dosadíme ze vztahu (33) směrodatnou odchylku $\sigma\{\epsilon_{t+h}(t)\}$ a vzniká hledaný interval

$$\left(\hat{y}_{t+h}(t) - 2\sqrt{1 + \sum_{j=1}^{h-1} \psi_j^2 \sigma_\epsilon}; \hat{y}_{t+h}(t) + 2\sqrt{1 + \sum_{j=1}^{h-1} \psi_j^2 \sigma_\epsilon} \right). \quad (36)$$

Do předpovědního intervalu typu (35), nebo (36) dosazujeme odhadnuté hodnoty parametrů modelu $ARMA(p, q)$. Vzorce nejsou moc citlivé na nepřesnosti, které vznikají při odhadování parametrů modelu $ARMA(p, q)$. Předpovědi v modelech $AR(p)$, $MA(q)$, $ARIMA(p, d, q)$ a $SARIMA(p, d, q) \times (P, D, Q)_L$ se sestavují obdobným způsobem, jaký jsme si popsali v této podkapitole.

3. Praktická část

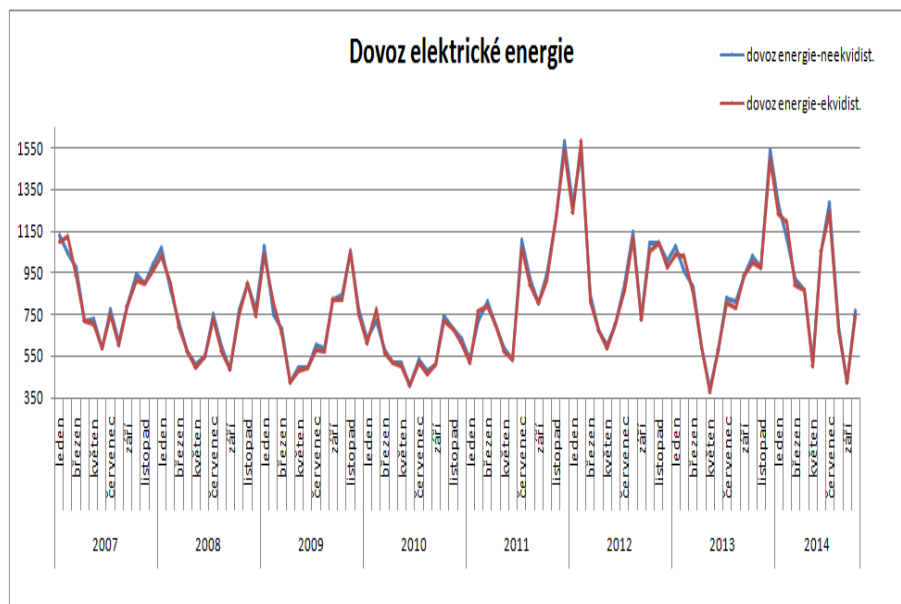
Teorii z předchozích kapitol si v této části aplikujeme na reálná data. Jde o veřejně dostupná data publikovaná na internetových stránkách Energetického regulačního úřadu, viz [7]. Data pochází z období od ledna 2007 do října 2014. Analyzováno je 94 měsíčních pozorování dovozu elektrické energie do České republiky, ztrát elektrické energie v distribučních sítích a výroby elektrické energie na území České republiky. Aktuální hodnoty od listopadu 2014 do června 2015 těchto dat si zobrazíme spolu s predikcemi a využijeme je pro srovnání predikovaných hodnot se skutečnými. Toto srovnání je nejlepším kritériem pro hodnocení kvality předpovědí.

3.1. Příprava dat, přípravné práce analytika

Nejdříve si musíme zpracovávaná data představit. Elektrická energie se nedá skladovat, ale na energetickém trhu jí musí být stále dostatek. Proto, i když je Česká republika významným vývozcem elektrické energie, musíme elektrickou energii v době největší spotřeby dovážet. Prvními zkoumanými daty byly právě data dovozu elektrické energie. Dále jsme analyzovali vývoj dat ztrát elektrické energie v distribučních sítích. Poslední analyzovaná data byly údaje o netto výrobě elektrické energie, kterou chápeme jako hrubou (brutto) výrobu elektřiny zmenšenou o náklady spojené s vlastní spotřebou elektřiny pro její výrobu (např. provoz reaktorů v jaderných elektrárnách, apod.). Hodnoty netto výroby elektřiny jsme získali prostým součtem výroby elektrické energie v následujících kategoriích

- AOE - výroba elektřiny pomocí alternativních elektráren např. výroba elektrické energie z biomasy, energie přílivu a příboje oceánů atd.,
- GOE - výroba elektřiny pomocí geotermálních elektráren. Jde o výrobu elektrické energie pomocí tepelných čerpadel. Tepelná čerpadla umožňují odnímat teplo z okolního prostředí, převádějí ho na vyšší teplotní hladinu a dále předávají k dalšímu využití, např. pro ohřev teplé užitkové vody,
- JE - výroba elektrické energie pomocí jaderných elektráren,

- PE - zahrnují výrobu elektrické energie s využitím parních elektráren,
- PPE - výroba elektrické energie paroplynovými elektrárnami,
- PSE - je energie vyrobená kombinací plynových a spalovacích elektráren,
- PVE - výroba elektrické energie z přečerpávacích vodních elektráren,
- VTE - jde o výrobu elektrické energie z větrných elektráren,
- SLE - výroba elektrické energie pomocí solárních panelů,
- VE - označuje výrobu elektrické energie pomocí vodních elektráren.



Obrázek 3: Dovoz elektrické energie

Data jsme nejdříve stáhli z internetových stránek ERÚ [7] a zapsali je do tabulky v programu Microsoft Excel. Protože data představují součet naměřených hodnot za každý den v příslušném kalendářním měsíci, vzniká nesoulad v délce pozorovaných časových úseků. Jedná se tedy o nekvidistantní časovou řadu, která je krátkodobá a intervalová. Abychom s touto řadou mohli odpovídajícím způsobem pracovat, museli jsme data převést na ekvidistantní časovou řadu. Využili

jsme k tomu vzorec $y_t^{(o)} = y_t \frac{\bar{d}}{d_m}$, kde \bar{d} je průměrný počet dní v měsíci, v našem případě zaokrouhleně 30, d_m je skutečný počet dní v m -tém kalendářním měsíci a $m = 1, 2, \dots, 12$. Provedli jsme tedy očišťování dat od jejich kalendářních variací a pro srovnání, k jakým došlo změnám, jsme si zobrazili data dovozu na obrázku 3. Všechny tři zkoumané časové řady jsme převedli na ekvidistantní. Grafy ostatních dat lze nalézt na příloženém CD. Soubor s daty, který byl vytvořen v MS Excel, jsme převedli do textového souboru *.txt. Pomocí příkazu `read.table('*.txt', sep = " \ ", header = TRUE)` jsme data načetli do programu *R*, ve kterém jsme již prováděli samotnou analýzu. Algoritmy použité pro analýzu dat a jejich zdrojové kódy taktéž naleznete na příloženém CD.

3.2. Aplikace spektrální analýzy

Časové řady jsme se rozhodli zkoumat pomocí aditivní dekompozice. Časovou řadu pak zapíšeme

$$y_t = T_t + C_t + S_t + \varepsilon_t,$$

kde T_t představuje trend, C_t a S_t tvoří dohromady periodickou složku časové řady a ε_t je náhodná složka. Spektrální analýza zkoumá periodickou složku časové řady. Před jejím určováním musíme ověřit, zda časová řada neobsahuje trend. Pokud by časová řada trend obsahovala, tak před aplikací spektrální analýzy musíme časovou řadu od trendu očistit.

3.2.1. Očištění dat od trendu

Trend v časových řadách jsme určovali v programu *R* pomocí následujících příkazů:

```
quadR3=lm(data ~ poly(t,3,raw=TRUE))
summary(quadR3)
```

První z těchto příkazů jsme využili pro určení koeficientů polynomu popisujícího trend zkoumané časové řady. Druhý příkaz jsme použili pro vyjádření základních charakteristik určených koeficientů. Celý algoritmus pro určení správného

řádu polynomu je uveden na příloženém CD, kde jsou dostupné veškeré zdrojové kódy vytvořené v programu *R*. V těchto algoritmech neuvádíme exponenciální a logistický trend, protože jejich užití pro zkoumaná data nebylo vhodné. Predikci pro jednotlivé trendy jsme získali užitím funkce *predict()*. Přesný popis jejího užití je uveden na CD.

Trend v datech dovozu elektrické energie

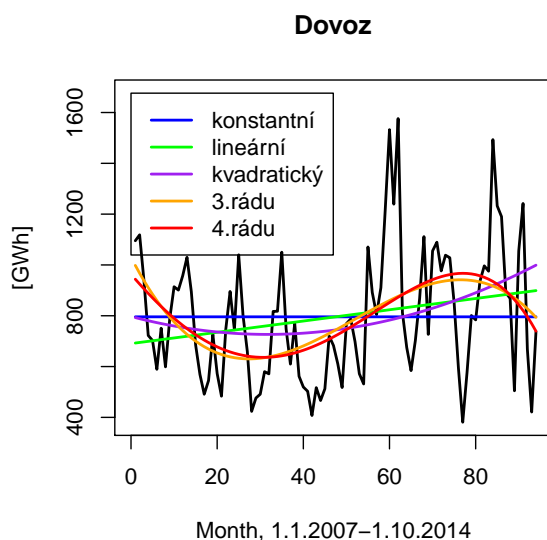
V programu *R* jsme aplikovali funkci *lm()* na data dovozu elektrické energie. V datech jsme se snažili odhalit trendovou funkci. Výsledky našeho zkoumání jsou uvedeny v tabulce 5. V prvním sloupci jsou uvedeny jednotlivé trendové funkce, ve druhém hodnoty modifikovaných indexů determinace R_M^2 . Použili jsme místo indexu determinace jeho modifikaci, protože odstraňuje závislost indexu determinace na počtu parametrů. Další sloupec obsahuje hodnoty *RMSE*, které představují druhou odmocninu hodnot *MSE*. V posledním sloupci jsou hodnoty *NRMSE*, které jsou dány vztahem (1). Jde o bezrozměrný ukazatel, který se užívá k vzájemnému porovnání časových řad.

Tabulka 5: Určování trendu dovoz

Trend	Trendová funkce y_t	R_M^2	<i>RMSE</i>	<i>NRMSE</i>
konstantní	795.80	-	255.4464	0.2137
lineární	$690.413 + 2.219t$	0.0453	248.2507	0.2077
kvadratický	$797.1027 - 4.449t + 0.070t^2$	0.0682	243.9113	0.2041
polynomický 3.řádu	$1030.524 - 33.177t + 0.822t^2 +$ $-0.005t^3$	0.1664	229.4385	0.1920

Hodnoty modifikovaných indexů determinace jsou velice nízké, ale my jsme se snažili pouze zachytit dlouhodobou tendenci vývoje této časové řady. Nechtěli jsme řadu modelovat jen trendovou funkcí, ale očistit ji pro následné užití spektrální analýzy. Určení správného trendu by bylo obtížné při rozhodování pouze na základě hodnot uvedených v tabulce 5. Proto jsme si data proložili vybranými trendy a jejich adekvátnost posuzovali graficky. Na obrázku 4 jsou zobrazena data proložená konstantním, lineárním, kvadratickým a polynomickým trendem tře-

tího a čtvrtého řádu. Při zkoumání grafu jsme zjistili, že data nevhodně prokládají lineární, kvadratický i polynomický trend čtvrtého řádu. Lineární a kvadratický trend mají odlišné tendence dlouhodobého vývoje než data dovozu. Naproti tomu polynomický trend čtvrtého řádu je vhodný, ale dosahuje obdobných výsledků jako polynomický trend třetího řádu, došlo by tak ke zbytečnému zesložiténí hledaného modelu.



Obrázek 4: Dovoz proložený trendy

```

Estimate Std. Error t value Pr(>|t|)
(Intercept) 795.80 26.49 30.04 <2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 256.8 on 93 degrees of freedom

```

Obrázek 5: Charakteristiky parametru konstantního trendu dovozu

K posuzování vhodnosti užití ostatních trendů jsme využívali testování významnosti parametrů. Pro provedení jednotlivých testů jsme využili hodnoty, které nám vypočítala funkce *summary()* v programu *R*. Výstupy po užití funkce


```

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    1.031e+03  1.007e+02  10.230 < 2e-16 ***
poly(t, 3, raw = TRUE)1 -3.318e+01  9.134e+00  -3.632 0.000467 ***
poly(t, 3, raw = TRUE)2  8.222e-01  2.228e-01   3.691 0.000383 ***
poly(t, 3, raw = TRUE)3 -5.277e-03  1.542e-03  -3.422 0.000936 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 234.5 on 90 degrees of freedom
Multiple R-squared:  0.1933,    Adjusted R-squared:  0.1664
F-statistic: 7.187 on 3 and 90 DF,  p-value: 0.0002225

```

Obrázek 6: Charakteristiky parametrů polynomického trendu 3. řádu dovozu

`summary()` jsou pro konstantní a polynomický trend třetího řádu zobrazeny na obrázcích 5 a 6. V prvním sloupci (Estimate) jsou hodnoty parametrů trendové funkce. Ve druhém sloupci (Std. Error) jsou uvedeny směrodatné odchylky těchto parametrů. Důležitým údajem pro určení významnosti parametrů je hodnota p -value, která vyjadřuje největší hladinu testu na níž nulovou hypotézu o významnosti parametrů ještě nezamítneme. Tuto hodnotu nalezneme ve čtvrtém sloupci ($Pr(> |t|)$). Parametr konstantního trendu i všechny parametry polynomického trendu třetího řádu jsme na základě hodnot p -value označili za významné. Oba zmíněné trendy se zdají být vhodné pro další analýzu. Vytvořili jsme tedy model jak pro konstantní, tak pro polynomický trend třetího řádu.

V případě konstantního trendu v datech dovozu nám vznikl odhad trendu ve tvaru

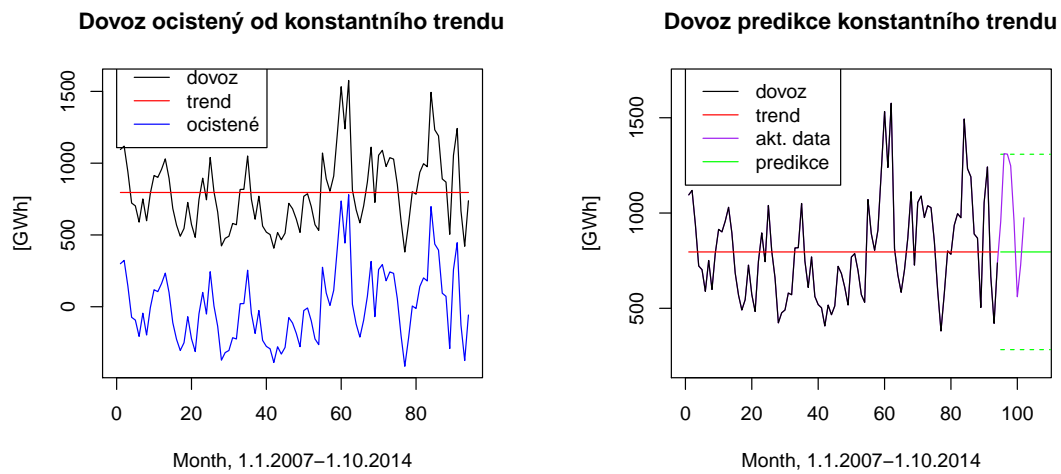
$$\hat{T}_t = 795.80.$$

Data jsme proložili konstantním trendem a následně je od něj očistili. To zobrazuje obrázek 7. Bodové i intervalové predikce získaného modelu jsou zobrazeny na obrázku 7 vpravo.

Polynomický trend třetího řádu dat dovozu je tvaru

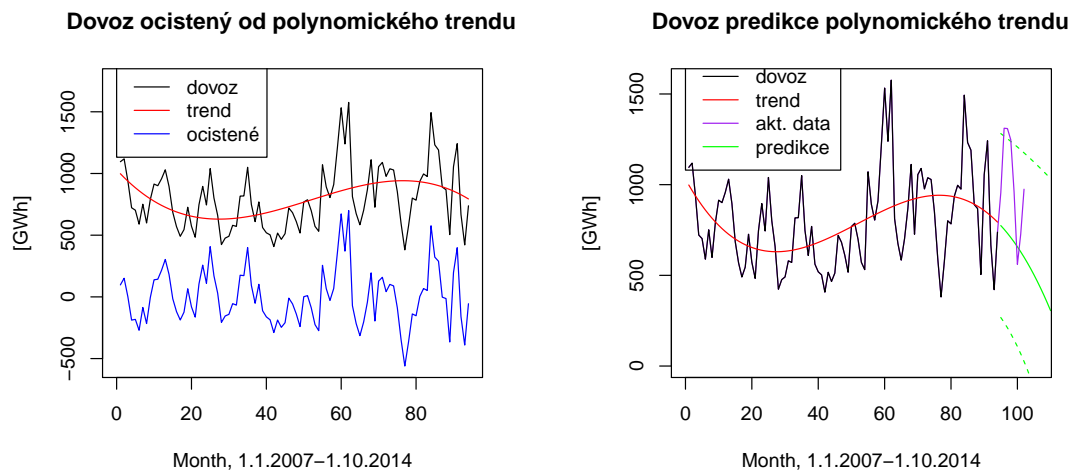
$$\hat{T}_t = 1030.524 - 33.177t + 0.822t^2 - 0.005t^3.$$

Na obrázku 8 jsou zobrazena data dovozu proložená polynomickým trendem třetího řádu, jsou zde i data očištěná od trendu (vlevo dole) a nakonec jsou na grafu



Obrázek 7: Dovoz očištěn od konstantního trendu a jeho predikce

vpravo zobrazeny bodové a intervalové predikce vycházející z uvedeného trendu.



Obrázek 8: Dovoz proložen polynomickým trendem 3. řádu a jeho predikce

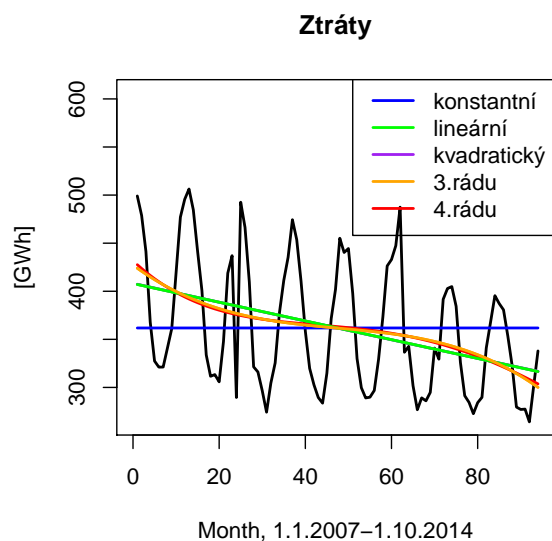
Trend v datech ztrát elektrické energie v sítích

Dále jsme zkoumali data ztrát elektrické energie v distribučních sítích. Při určování trendové složky modelu jsme postupovali obdobně jako v předešlé části. Všechny uvažované trendové funkce pro data ztrát jsou zapsány v tabulce 6.

Hodnoty modifikovaných indexů determinace jsou velmi nízké, a proto jsme opět vycházeli z grafického zobrazení, viz obrázek 9. Na obrázku není zobrazen kvadratický trend a polynomický trend čtvrtého řádu. Křivka kvadratického trendu proložená původními daty je totožná s křivkou lineárního trendu. Křivka polynomického trendu čtvrtého řádu je téměř překryta křivkou polynomického trendu třetího řádu.

Tabulka 6: Určování trendu ztráty

Druh trendu	Trendová funkce y_t	R_M^2	$RMSE$	$NRMSE$
konstantní	361.815	-	67.4867	0.2787
lineární	$408.073 - 0.974t$	0.1441	62.0983	0.2565
kvadratický	$408.301 - 0.988t + 0.0002t^2$	0.1347	62.0982	0.2565
polynomický 3.řádu	$427.234 - 3.318t + 0.061t^2 +$ $-0.0004t^3$	0.1353	61.7342	0.2550



Obrázek 9: Ztráty proložené trendy

Z grafu na obrázku 9 je dále patrné, že křivka polynomického trendu třetího řádu se od křivky lineárního trendu odlišuje jen minimálně. Proložení dat křivkou

polynomického trendu třetího řádu vedlo k užití příliš mnoha parametrů. Vzhledem k povaze dat je zřejmé, že nejvhodnější byla volba lineárního trendu. Ten nejlépe odpovídá hodnotám původní časové řady.

```

                Estimate Std. Error t value Pr(>|t|)
(Intercept) 408.0732    13.0524  31.264 < 2e-16 ***
t           -0.9739     0.2386  -4.082 9.52e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

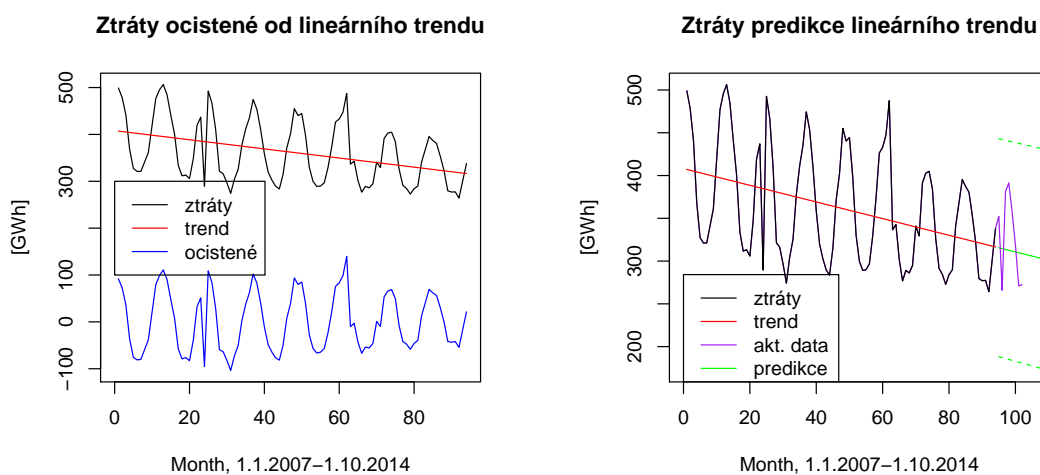
Residual standard error: 62.77 on 92 degrees of freedom
Multiple R-squared:  0.1533,    Adjusted R-squared:  0.1441
F-statistic: 16.66 on 1 and 92 DF,  p-value: 9.523e-05

```

Obrázek 10: Charakteristiky parametrů lineárního trendu dat ztrát

Hodnoty parametrů lineárního trendu jsou dle p-value významné, to zobrazuje poslední sloupec na obrázku 10. Lineární trend dat ztrát je tvaru

$$\hat{T}_t = 408.0732 - 0.9739t.$$



Obrázek 11: Ztráty proložené lineárním trendem a jejich predikce

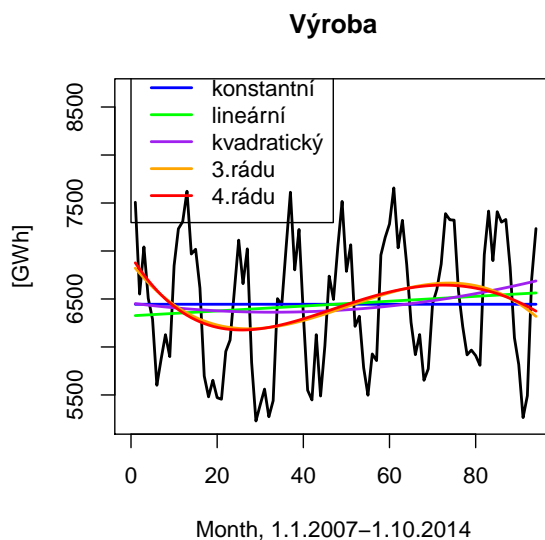
Na obrázku 11 vlevo je zobrazeno proložení dat ztrát lineárním trendem a jejich očištění od trendu. Graf vpravo zobrazuje bodové a intervalové predikce hodnot časové řady vycházející z lineárního trendu.

Trend v datech výroby elektrické energie

Poslední zkoumanou časovou řadou byla data výroby elektrické energie. Jednotlivé trendové funkce, které jsme využili v případě dat výroby elektrické energie, jsou uvedeny v tabulce 7. Data výroby jsme proložili těmito trendy, což je zobrazeno na obrázku 12.

Tabulka 7: Určování trendu výroby

Typ trendu	Trendová funkce y_t	R_M^2	$RMSE$	$NRMSE$
konstantní	6444.952	-	688.5086	0.2835
lineární	$6324.355 + 2.539t$	0.0007	685.0535	0.2821
kvadratický	$6457.522 - 5.784t + 0.087t^2$	0.0046	682.6206	0.2811
polynomický 3.řádu	$6877.435 - 57.463t + 1.440t^2 +$ $-0.009t^3$	0.0326	666.1837	0.2743



Obrázek 12: Výroba proložená trendy

Křivky lineárního i kvadratického trendu mají jiné tendence dlouhodobého vývoje než zkoumaná data. Křivka polynomického trendu čtvrtého řádu je identická s křivkou polynomického trendu třetího řádu. Křivka polynomického trendu

třetího řádu prokládá zkoumaná data vhodně, ale vedla k přílišnému zesložnění modelu. Upřednostnili jsme zde nakonec konstantní trend. Významnost jeho parametru je ověřena v posledním sloupci na obrázku 13.

```

                Estimate Std. Error t value Pr(>|t|)
(Intercept)    6444.9      71.4    90.27  <2e-16 ***

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 692.2 on 93 degrees of freedom

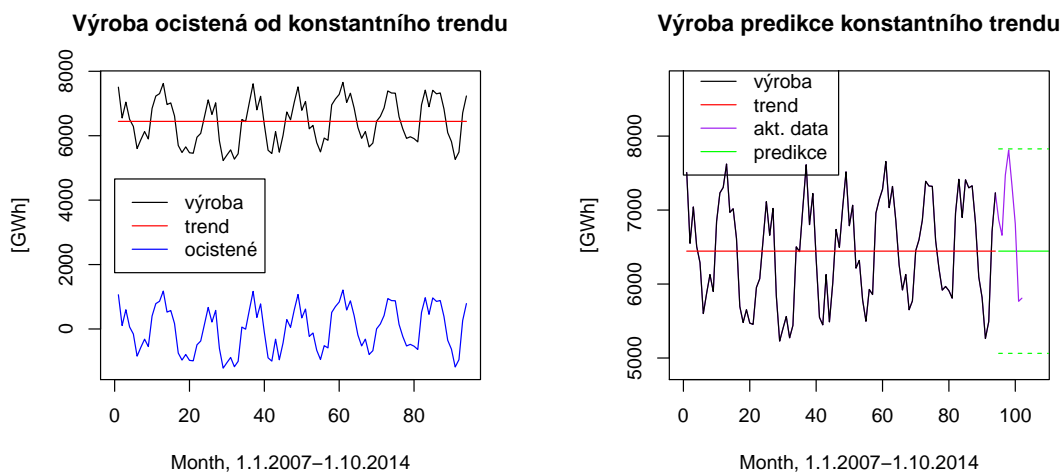
```

Obrázek 13: Významnost parametru konstantního trendu výroby

Konstantní trend dat výroby je tvaru

$$\hat{T}_t = 6444.952.$$

Na obrázku 14 jsou zobrazena data proložená konstantním trendem, v dolní části je zobrazena časová řada již očištěná od zvoleného konstantního trendu a nakonec jsou v pravé části obrázku 14 zobrazeny bodové i intervalové predikce hodnot časové řady vycházející z konstantního trendu.



Obrázek 14: Konstantní trend a predikce výroby

3.2.2. Spektrální analýza

Pro potřeby aplikace spektrální analýzy jsme zkoumaná data museli očistit od trendové složky, což lze vyjádřit jako

$$y_t^{(0)} = y_t - \hat{T}_t.$$

Snažili jsme se odhalit významné periody. K jejich hledání jsme využili algoritmus v programu *R*, jehož zdrojové kódy lze nalézt na přiloženém CD. V algoritmu jsme pro nalezení významných period použili Fisherův test včetně jeho modifikace v podobě Siegelova testu. Siegelův test se užívá při složené periodicitě, neboť Fisherův test v takovém případě nedává příliš kvalitní výsledky. Siegelův test ověří významnost period, které jsme našli pomocí Fisherova testu. Po provedení obou testů na významnost period jsme vytvořili pro všechna zkoumaná data model skrytých period a následně bodovou i intervalovou predikci hodnot časové řady.

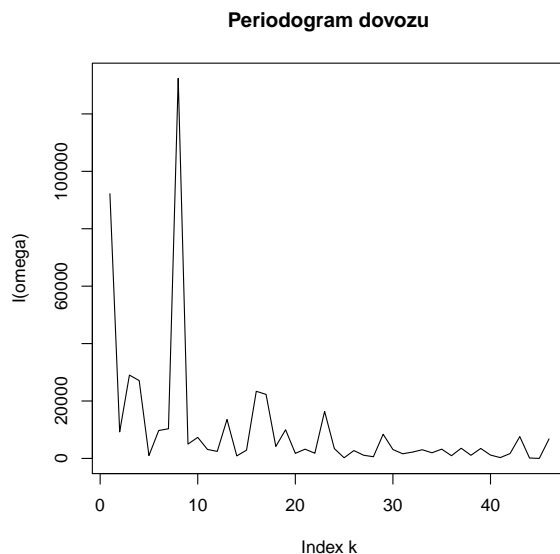
3.2.3. Spektrální analýza dat dovozu elektrické energie

Při hledání nejvhodnějšího trendu jsme vybrali jako významné trendové funkce dat dovozu konstantní a polynomický trend třetího řádu. Z těchto důvodů jsme museli spektrální analýzu provést zvlášť pro oba modely. Celý postup spektrální analýzy jsme provedli nejdříve pro model s konstantním trendem a poté jsme ho zopakovali i pro model s polynomickým trendem třetího řádu.

Spektrální analýza dovozu v modelu s konstantním trendem

Nejdříve jsme data dovozu elektrické energie očistili od konstantního trendu. Poté jsme z očištěných dat dovozu vypočítali hodnoty periodogramu daného vztahem (6). Hodnoty periodogramu, které jsme získali z algoritmu v programu *R*, jsou zobrazeny na obázku 15.

Z grafu je patrné, že časová řada může obsahovat hned několik významných period. Následně jsme hodnoty periodogramu normovali pomocí vzorce (7), normované hodnoty jsme poté testovali pomocí Fisherova testu periodicity a určovali



Obrázek 15: Periodogram dat dovozu očištěných od konstantního trendu

jejich významnost. Ověřovali jsme, zda platí hypotéza

$$H_0 : y_t^{(0)} = \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma_\varepsilon^2),$$

která označuje testovanou periodu za nevýznamnou. Z normovaných hodnot jsme vybrali největší a získali jsme hodnotu testové statistiky

$$V = \frac{I(\omega_8)}{\sum_{j=1}^{46} I(\omega_j)} = \frac{132383.2}{487713.5} = 0.2714.$$

Tuto hodnotu jsme porovnávali s kritickou hodnotou $g_F = 0.141$ Fisherova testu na hladině významnosti $\alpha = 0.05$ pro $m = 46$, která je uvedena v [13, str. 12]. Protože platí $V > g_F(0.05)$, zamítli jsme nulovou hypotézu. Fisherův test nám na hladině významnosti 5% určil jako významnou frekvenci $\omega_8 = 0.5347$. Délku periody odpovídající frekvenci ω_8 jsme určili ze vztahu

$$\frac{2\pi}{\omega_j} = \frac{2\pi}{\frac{2\pi j}{n}} = \frac{2\pi n}{2\pi j} = \frac{94}{8} = 11.75.$$

Délka periody je tedy 11 měsíců a 22.5 dnů, neboť v případě ekvidistantních časových řad uvažujeme všechny měsíce stejně dlouhé, tedy 30 dní. Pokračovali

jsme v testování, testovanou hodnotu $I(\omega_8)$ jsme vynechali, položili jsme $m = 45$ a Fisherův test zopakovali pro novou největší hodnotu. Novou největší hodnotu jsme porovnali s kritickou hodnotou Fisherova testu pro $m = 45$ na hladině významnosti $\alpha = 0.05$. Testování významnosti period jsme prováděli, dokud jsme pro nějakou z testovaných hodnot nulovou hypotézu nezamítli a poslední testovanou hodnotu tedy označili za nevýznamnou. Výsledky testování nám shrnuje tabulka 8.

Tabulka 8: Fisherův test u dat dovozu očištěných od konstantního trendu

m	V	j	$g_F(0.05)$	$g_F(0.01)$	na 5% hladině	na 1% hladině
46	0.2714	8	0.141	0.184	zamítám H_0	zamítám H_0
45	0.2595	1	0.143	0.188	zamítám H_0	zamítám H_0
44	0.1104	3	0.146	0.192	nezamítám H_0	nezamítám H_0

Pomocí Fisherova testu jsme na hladině významnosti $\alpha = 0.05$ i $\alpha = 0.01$ určili jako významné periody příslušející úhlové frekvenci $\omega_8 = 0.5347$ a frekvenci $\omega_1 = 0.0668$. Frekvence ω_1 je ekvivalentní délce periody 94 měsíců.

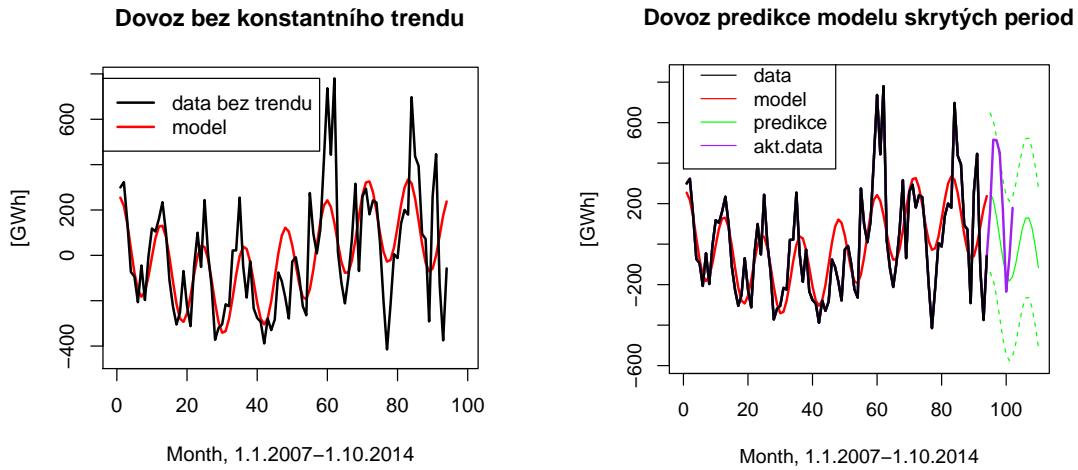
Pro vytvoření modelu skrytých period jsme museli určit jeho parametry. Odhadnutou hodnotu parametru $\hat{\mu} = 0.1201 \cdot 10^{-12}$ jsme získali ze vztahu (3). Z důvodu její zanedbatelnosti jsme ji v modelu neuváděli. Odhadnutou hodnotu parametru $\hat{\alpha}$ jsme určili ze vztahu (4) a odhad parametru $\hat{\beta}$ získali aplikací vztahu (5). Přesnost získaných odhadů parametrů je uvedena v tabulce 9. Model skrytých period dat dovozu je ve tvaru

$$\begin{aligned} \hat{y}_t^{(0)} &= \hat{\alpha}_1 \cos(t\omega_1) + \hat{\beta}_1 \sin(t\omega_1) + \hat{\alpha}_8 \cos(t\omega_8) + \hat{\beta}_8 \sin(t\omega_8) = \\ &= 162.0834 \cos(0.0668t) + 95.5208 \sin(0.0668t) + \\ &+ 75.6021 \cos(0.5347t) - 137.6084 \sin(0.5347t). \end{aligned}$$

Model je zobrazen na obrázku 16. Vlevo jsou data dovozu očištěná od konstantního trendu proložená určeným modelem skrytých period. Graf umístěný vpravo zobrazuje bodové a intervalové predikce dat vytvořené z nalezeného modelu. Modifikovaný index determinace modelu je 0.4358. Další ukazatelé jsou $RMSE = 187.6949$ a $NRMSE = 0.1570$.

Tabulka 9: Přesnost odhadů parametrů modelu skrytých period u dovozu očištěného od konstantního trendu

Parametr	$\hat{\alpha}_1$	$\hat{\alpha}_8$	$\hat{\beta}_1$	$\hat{\beta}_8$
Odhad	162.0834	75.6021	95.5208	-137.6084
Směrodatná odchylka	28.14	28.14	28.14	28.14



Obrázek 16: Model skrytých period a jeho predikce u dat dovozu očištěných od konstantního trendu

Fisherovým testem jsme našli v datech očištěných od konstantního trendu dvě významné periody. Protože Siegelův test dává kvalitnější výsledky při složené periodicitě než Fisherův test, bylo vhodné si jejich významnost ověřit. Všechny hodnoty periodogramu jsme normovali. Jako významné periody jsme označili pouze ty, které zůstaly zachovány v součtu (10). tj. hodnoty, pro které platí $(Y_j - \lambda g_F) > 0$. Konstanta λ je rovna 0.6 a hodnoty g_F představují kritické hodnoty Fisherova testu. Na hladině významnosti 5% pro $\lambda = 0.6$ a $m = 46$ platí

$$T_{0.6} = \sum_{j=1}^{46} (V_j - 0.6g_F)^+ = 0.2913 > 0.0653.$$

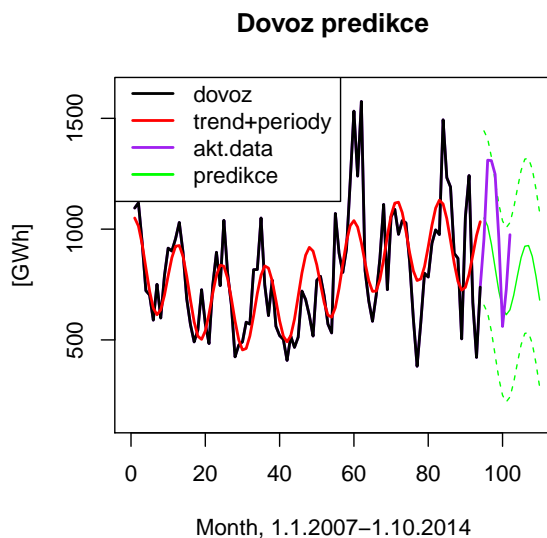
Siegelův test na hladině významnosti 5% nulovou hypotézu zamítá. Ve sčítanci zůstaly zachovány hodnoty s indexy $j = 8$ a $j = 1$. Do testové statistiky přispěly dva členy součtu a Siegelův test určuje jako významné periody s frekvencí ω_8 a ω_1 .

Siegelův test nám potvrdil závěry získané pomocí Fisherova testu. Obě určené periody jsou významné.

Pro data dovozu jsme našli model tvaru

$$\hat{y}_t = 795.80 + 162.0834 \cos(0.0668t) + 95.5208 \sin(0.0668t) + \\ + 75.6021 \cos(0.5347t) - 137.6084 \sin(0.5347t).$$

Data dovozu elektrické energie jsou zobrazeny s konstantním trendem, modelem skrytých period a predikcí na obrázku 17. Hodnoty bodové a intervalové predikce jsou uvedeny v tabulce 10. Modifikovaný index determinace je 0.4358. Další ukazatelé modelu jsou $RMSE = 187.6949$ a $NRMSE = 0.1570$. Graf na obrázku 17 je doplněn o aktuální data dovozu, která jsou zobrazena fialovou barvou. Bodové predikce kopírují dlouhodobou tendenci vývoje dat dovozu. Data náleží do relativně úzkých predikčních intervalů, a proto jsme nalezený model hodnotili jako kvalitní. Model adekvátně vystihnul budoucí chování dat dovozu.



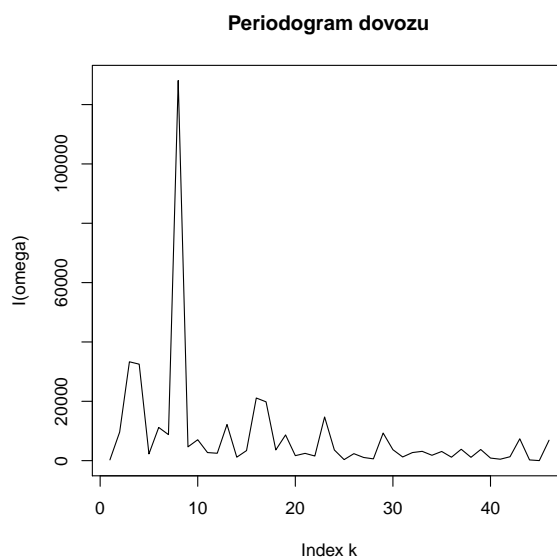
Obrázek 17: Dovoz s konstantním trendem, modelem skrytých period a predikcí

Tabulka 10: Predikce dat dovozu s konstantním trendem a modelem skrytých period

t	bodové	95% predikční intervaly		t	bodové	95% predikční intervaly	
95	1050.182	656.841	1443.522	99	720.502	327.162	1113.842
96	1014.051	620.711	1407.391	100	643.585	250.244	1036.925
97	932.533	539.193	1325.873	101	613.529	220.189	1006.869
98	825.664	432.324	1219.004	102	635.852	242.512	1029.192

Spektrální analýza dovozu v modelu s polynomickým trendem 3. řádu

Opět jsme nejprve očistili data od polynomického trendu třetího řádu. Dále jsme postupovali jako v předchozím případě. Z očištěných dat dovozu jsme vypočítali hodnoty periodogramu, které jsou znázorněny na obrázku 18. Normované hodnoty periodogramu jsme testovali na významné periody pomocí Fisherova testu periodicity. Výsledky testování na hladině významnosti 5% a 1% jsou uvedeny v tabulce 11.



Obrázek 18: Periodogram dovozu očištěného od polynomického trendu 3. řádu

Z tabulky plyne, že pomocí Fisherova testu na hladině významnosti 1% a 5% jsme určili jako významnou pouze jedinou frekvenci $\omega_8 = 0.5347$, která odpovídá

Tabulka 11: Fisherův test u dat dovozu očištěných od polynomického trendu třetího řádu

m	V	j	$g_F(0.05)$	$g_F(0.01)$	na 5% hladině	na 1% hladině
46	0.3259	8	0.141	0.184	zamítám H_0	zamítám H_0
45	0.1257	3	0.143	0.188	nezamítám H_0	nezamítám H_0

délce periody 11 měsíců a 22.5 dnů. Model skrytých period pro data dovozu očištěná od polynomického trendu třetího řádu je ve tvaru

$$\hat{y}_t^{(0)} = 163.4846 \cos(0.5347t) + 86.7345 \sin(0.5347t),$$

kde $\hat{\mu} = 0.3144 \cdot 10^{-12}$. Parametr $\hat{\mu}$ z důvodu zanedbatelnosti v modelu neuvádíme. Přesnost odhadů parametrů modelu skrytých period nalezneme v tabulce 12.

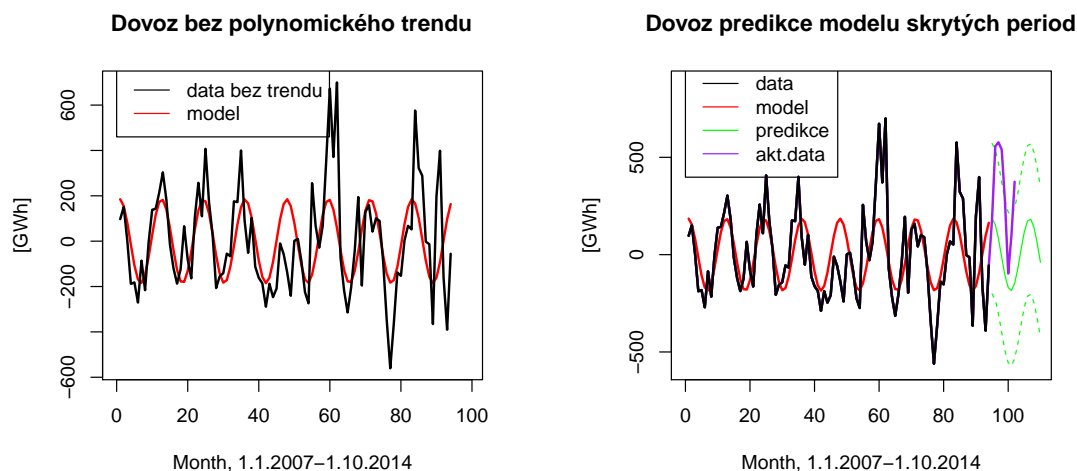
Tabulka 12: Přesnost odhadů parametrů modelu skrytých period podle Fisherova testu dat dovozu očištěných od polynomického trendu 3. řádu

Parametr	$\hat{\alpha}_8$	$\hat{\beta}_8$
Odhad	163.4846	86.7345
Směrodatná odchylka	27.94	27.94

Model skrytých period určený pomocí Fisherova testu je s daty dovozu očištěnými od polynomického trendu třetího řádu zobrazen na obrázku 19 vlevo. Na grafu vpravo jsou bodové a intervalové predikce hodnot časové řady z modelu skrytých period, dále jsou pro srovnání uvedena i aktuální data. Modifikovaný index determinace R_M^2 je 0.3105. Hodnoty sledovaných ukazatelů jsou $RMSE = 188.4595$ a $NRMSE = 0.1495$.

Pomocí Fisherova testu jsme našli pouze jednu významnou frekvenci. Protože hodnota testové statistiky odpovídající frekvenci ω_3 je velice blízko kritické hodnotě Fisherova testu, můžeme vyslovit domněnku o významnosti i této frekvence. Domněnku jsme si ověřili pomocí Siegelova testu. Testová statistika Siegelova testu na hladině významnosti 5% pro $\lambda = 0.6$ je

$$T_{0.6} = \sum_{j=1}^{46} (V_j - 0.6g_F)^+ = 0.2413 > 0.0653.$$



Obrázek 19: Model skrytých period určený Fisherovým testem dat dovozu bez polynomického trendu 3. řádu a jeho predikce

Ve sčítanci zůstávají zachovány hodnoty s indexy $j = 8$ a $j = 3$. Do testové statistiky přispěly dva členy součtu a Siegelův test určuje jako významné periody o frekvencích $\omega_8 = 0.5347$ a $\omega_3 = 0.2005$. Frekvence ω_3 odpovídá délce periody 2 roky a 7.33 měsíců.

Dále jsme vytvořili model skrytých period určený Siegelovým testem pro data dovozu elektrické energie očištěná od polynomického trendu třetího řádu, který je ve tvaru

$$\hat{y}_t^{(0)} = 69.4041 \cos(0.2005t) - 63.9603 \sin(0.2005t) + \\ + 163.4846 \cos(0.5347t) + 86.7345 \sin(0.5347t).$$

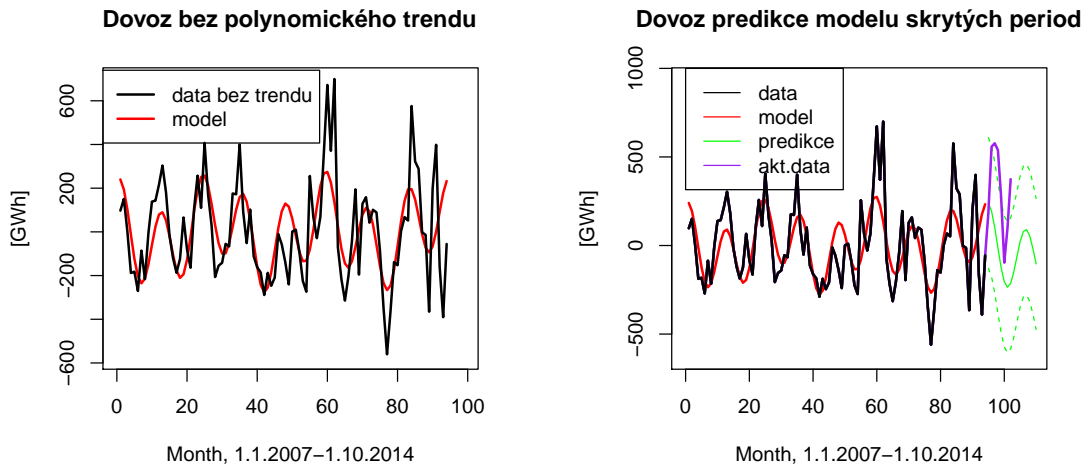
Odhadnutá hodnota $\hat{\mu}$ je opět zanedbatelná $0.3144 \cdot 10^{-12}$. Přesnost ostatních odhadů parametrů je uvedena v tabulce 13. Modifikovaný index determinace je 0.3834 a další ukazatelé modelu jsou $RMSE = 176.2472$ a $NRMSSE = 0.1399$.

Na obrázku 20 je zobrazen tento model s daty očištěnými od polynomického trendu třetího řádu. Vpravo jsou opět uvedeny bodové a intervalové predikce hodnot časové řady.

Časová řada dat dovozu s polynomickým trendem třetího řádu a modelem skrytých period určených Siegelovým testem je zobrazena na obrázku 21. Vý-

Tabulka 13: Přesnost odhadů parametrů modelu skrytých period podle Siegelova testu u dat dovozu očištěných od polynomického trendu 3. řádu

Parametr	$\hat{\alpha}_3$	$\hat{\alpha}_8$	$\hat{\beta}_3$	$\hat{\beta}_8$
Odhad	69.4041	163.4846	-63.9603	86.7345
Směrodatná odchylka	26.42	26.42	26.42	26.42

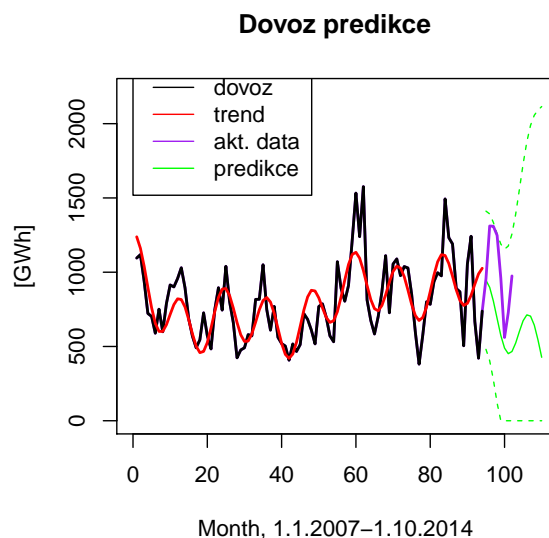


Obrázek 20: Model skrytých period určený Siegelovým testem dat dovozu bez polynomického trendu 3. řádu a jeho predikce

sledný model dat dovozu je tvaru

$$\hat{y}_t = 1030.524 - 33.177t + 0.822t^2 - 0.005t^3 + 69.4041 \cos(0.2005t) + \\ - 63.9603 \sin(0.2005t) + 163.4846 \cos(0.5347t) + 86.7345 \sin(0.5347t).$$

Sledované charakteristiky modelu jsou $R_M^2 = 0.4852$, $RMSE = 176.2472$ a hodnota $NRMSE = 0.1475$. Graf na obrázku 21 je doplněn o aktuální data za období od listopadu 2014 do června 2015. Bodové predikce se z počátku odchýlily od zobrazených dat, ale v prostřední části se bodový odhad datům přiblížil. Predikční intervaly modelu se podstatně rozšířily v porovnání s předchozím modelem. Aktuální hodnoty časové řady se nachází uvnitř predikčních intervalů, a proto i model s polynomickým trendem třetího řádu a významnými frekvencemi ω_3 a ω_8 hodnotíme jako kvalitní. Hodnoty bodové predikce a meze predikčních intervalů jsou uvedeny v tabulce 14.



Obrázek 21: Dovoz s polynomičným trendem 3. řádu, modelem skrytých period určený Siegelovým testem a predikcí

Tabulka 14: Predikce dat dovozu s polynomičným trendem 3. řádu a modelem skrytých period

t	bodové	95% predikční intervaly		t	bodové	95% predikční intervaly	
95	1014.598	480.772	1411.229	99	560.489	0	1189.240
96	947.490	403.855	1392.448	100	453.896	0	1155.481
97	834.254	276.128	1334.718	101	394.334	0	1176.037
98	696.072	118.715	1257.957	102	386.619	0	1257.374

Srovnání nalezených modelů dat dovozu elektrické energie

Oba způsoby zpracování dat dovozu si nyní porovnáme. Při očištění dat od konstantního trendu jsme Fisherovým testem i jeho modifikací určili jako významné dvě periody. Jednalo se o periody příslušející frekvencím ω_8 a ω_1 . U dat dovozu očištěných od polynomičného trendu jsme Siegelovým testem určili také dvě významné frekvence, v tomto případě šlo o frekvence ω_8 a ω_3 . Ukazatelé modelů jsou shrnuty v tabulce 15. Vyšší hodnota modifikovaného indexu determinace a nižší hodnota ukazatelů $RMSE$ a $NRMSE$ svědčí pro model s polynomičným trendem třetího řádu. Model s polynomičným trendem třetího řádu je kvalitnější

pro zachycení dat, ale jeho predikce je méně vhodná než u modelu s konstantním trendem. Predikční intervaly modelu s polynomickým trendem třetího řádu jsou mnohem více neurčité. Pokud bychom chtěli data pouze zachytit, je lepší model s polynomickým trendem třetího řádu. Naším záměrem bylo data také predikovat, a proto jako vhodnější pro modelování dat dovozu jsme zvolili model s konstantním trendem.

Tabulka 15: Porovnání určených modelů u dat dovozu

Dovoz	konst. trend s periodami	polynom. trend 3. řádu s periodami
R_M^2	0.4358	0.4852
RMSE	187.6949	176.2472
NRMSE	0.1570	0.1475

3.2.4. Spektrální analýza dat ztrát elektrické energie v sítích

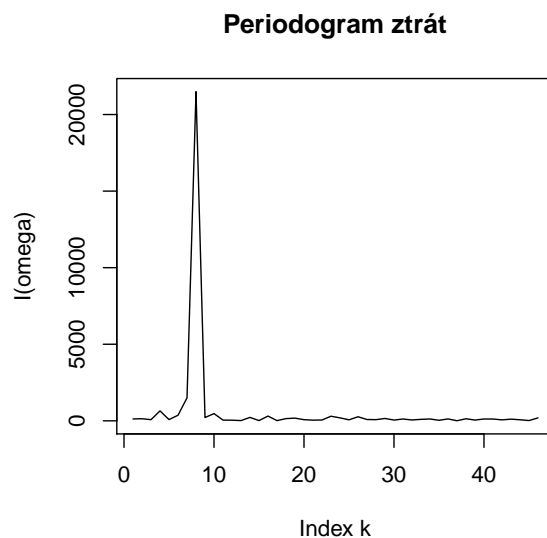
Zpracováním dat ztrát elektrické energie v distribučních sítích očištěných od lineárního trendu jsme získali pomocí (6) periodogram zobrazený na obrázku 22. V tabulce 16 je shrnut průběh testování významnosti period Fisherovým testem na hladinách významnosti 1% a 5%.

Tabulka 16: Fisherův test u dat ztrát očištěných od lineárního trendu

m	V	j	$g_F(0.05)$	$g_F(0.01)$	na 5% hladině	na 1% hladině
46	0.7451	8	0.141	0.184	zamítám H_0	zamítám H_0
45	0.2036	7	0.143	0.188	zamítám H_0	zamítám H_0
44	0.1107	4	0.146	0.192	nezamítám H_0	nezamítám H_0

Odhalili jsme významné frekvence $\omega_8 = 0.5347$ a $\omega_7 = 0.4679$. Frekvenci ω_8 odpovídá délka periody 11 měsíců a 22.5 dnů. Frekvence ω_7 je ekvivalentní délce periody 13 měsíců a 12.86 dnů. Při vytváření modelu skrytých period jsme opět vynechali zanedbatelný odhad parametru $\hat{\mu} = 0,2428 \cdot 10^{-13}$. Model skrytých period dat ztrát očištěných od lineárního trendu je tvaru

$$\hat{y}_t^{(0)} = 44.8413 \cos(0.5347t) + 61.1177 \sin(0.5347t) - 5.8072 \cos(0.4679t) + \\ - 19.1446 \sin(0.4679t).$$



Obrázek 22: Periodogram dat ztrát očištěných od lineárního trendu

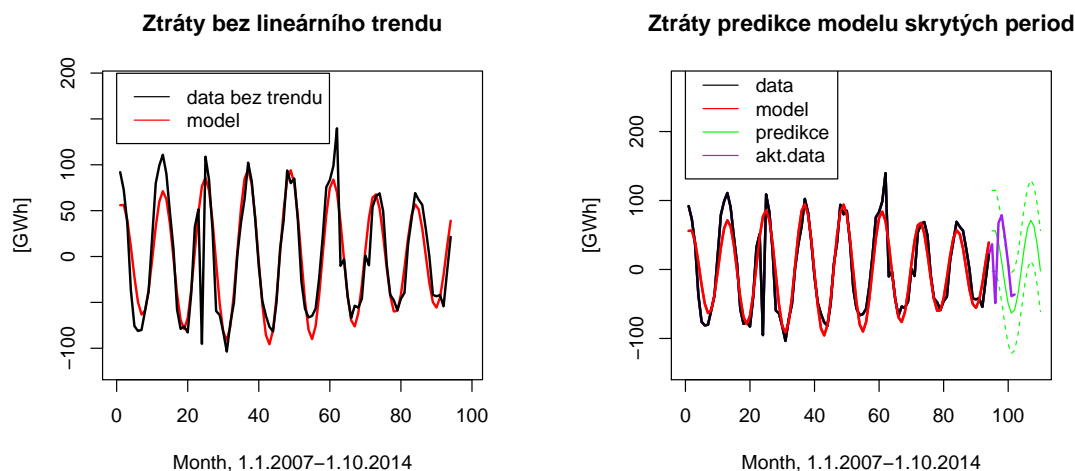
Přesnost odhadů parametrů modelu skrytých period je uvedena v tabulce 17. Model skrytých period dat očištěných od lineárního trendu je zobrazen na obrázku 23. Vpravo jsou bodové a intervalové predikce hodnot časové řady doplněné o aktuální data ztrát elektrické energie v sítích. Modifikovaný index determinace modelu je 0.7878. Další sledované charakteristiky jsou $RMSE = 27.9825$ a $NRMSE = 0.1150$.

Tabulka 17: Přesnost odhadů parametrů modelu skrytých period určených Fisherovým testem u dat ztrát očištěných od lineárního trendu

Parametr	$\hat{\alpha}_7$	$\hat{\alpha}_8$	$\hat{\beta}_7$	$\hat{\beta}_8$
Odhad	-5.8072	44.8413	-19.1446	61.1177
Šměrodatná odchylka	4.195	4.195	4.195	4.195

Významné periody jsme ověřili na hladině významnosti 5% Siegelovým testem pro $\lambda = 0.6$. Testovací statistika je ve tvaru

$$T_{0.6} = \sum_{j=1}^{46} (V_j - 0.6g_F)^+ = 0.6605 > 0.0653.$$



Obrázek 23: Model skrytých period určený Fisherovým testem dat ztrát očištěných od lineárního trendu a jeho predikce

Siegelův test nulovou hypotézu zamítá na hladině významnosti 5%. Do testové statistiky přispěl pouze sčítanec s indexem $j = 8$, a proto je model skrytých period tvaru

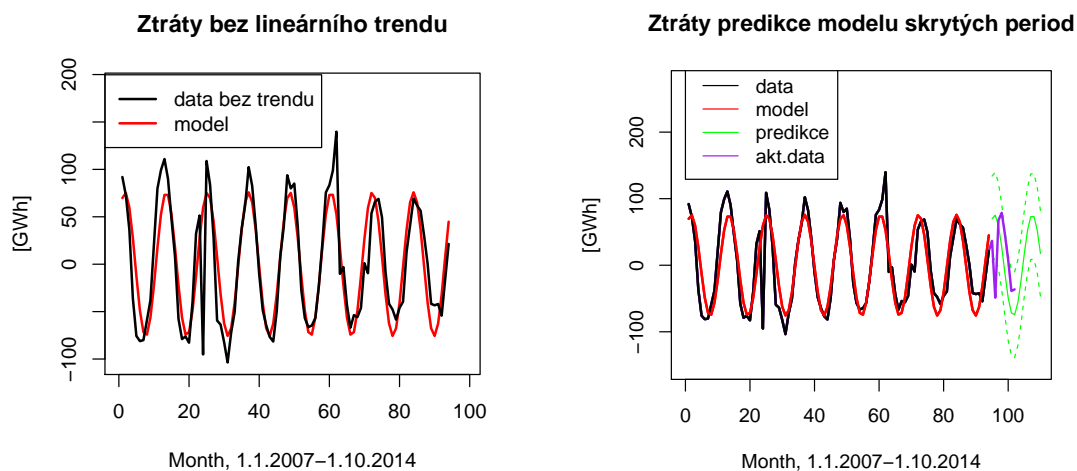
$$\hat{y}_t^{(0)} = 44.8413 \cos(0.5347t) + 61.1177 \sin(0.5347t). \quad (37)$$

Hodnota $\hat{\mu}$ je $0,2428 \cdot 10^{-13}$, odhad tohoto parametru je opět zanedbatelný. Přesnost odhadů parametrů modelu je uvedena v tabulce 18. Ukazatelé modelu jsou $R_M^2 = 0.7394$, $RMSE = 31.3551$ a $NRMSE = 0.1288$.

Tabulka 18: Přesnost odhadů parametrů modelu skrytých period určených Siegelovým testem dat ztrát očištěných od lineárního trendu

Parametr	$\hat{\alpha}_8$	$\hat{\beta}_8$
Odhad	44.8413	61.1177
Směrodatná odchylka	4.648	4.648

Model (37) prokládá data na grafu v obrázku 24 vlevo. V pravé části obrázku 24 je zobrazen model doplněný o intervalové a bodové predikce a aktuální data. Bodové predikce dat odpovídají tendenci skutečně naměřených hodnot. Aktuální data v některých bodech meze intervalové predikce mírně překročují.

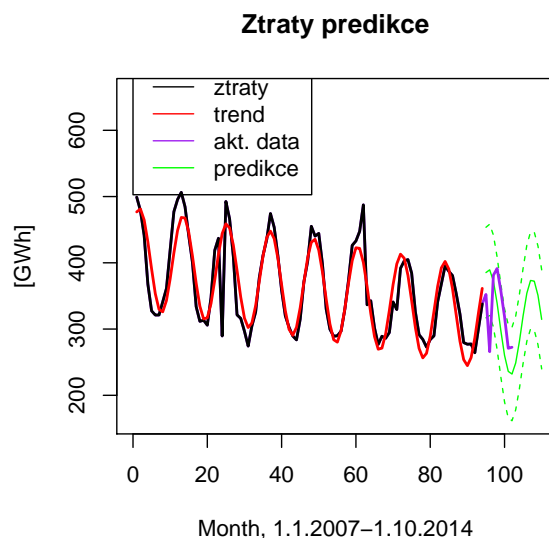


Obrázek 24: Model skrytých period určený Siegelovým testem dat ztrát očištěných od lineárního trendu a jeho predikce

Celkový námi nalezený model obsahující lineární trend a model skrytých period určený pomocí Siegelova testu, který popisuje data ztrát elektrické energie v sítích je ve tvaru

$$\hat{y}_t = 408.073 - 0.974t + 44.8413 \cos(0.5347t) + 61.1177 \sin(0.5347t).$$

Tento model je zobrazen na obrázku 25 včetně bodové a intervalové predikce a aktuálních dat. Bodové predikce kvalitně předpovídají budoucí tendenci dat. Intervalové predikce jsou relativně úzké, a proto by měl model kvalitně vystihnout budoucí vývoj dat. Aktuální data meze intervalové predikce mírně překročují, ale v koncové části kopírují meze intervalové predikce. Model vzhledem k šířce predikčních intervalů zachytil budoucí tendenci dat dostatečně kvalitně. V tabulce 19 jsou uvedeny bodové a intervalové predikce hodnot z celkového námi nalezeného modelu pro popis dat ztrát. Modifikovaný index determinace modelu s lineárním trendem a modelem skrytých period s jednou významnou frekvencí je 0.7769. Ukazatel $RMSE$ je 31.3551 a normovaný ukazatel $NRMSE$ dává hodnotu 0.1295.



Obrázek 25: Ztráty s lineárním trendem, modelem skrytých period určeným Siegelovým testem a predikcí

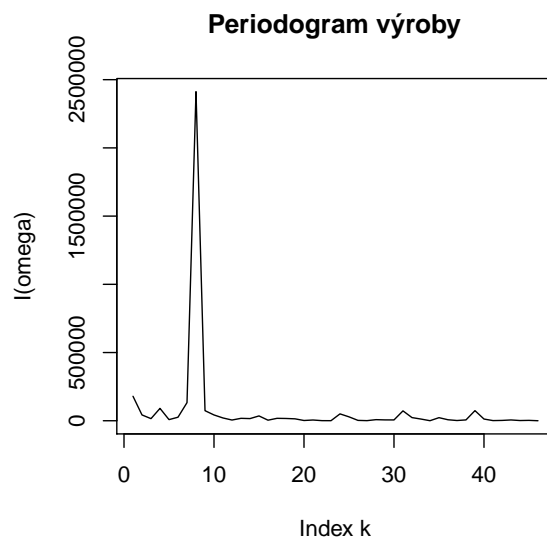
Tabulka 19: Predikce dat ztrát s lineárním trendem a modelem skrytých period

t	bodové	95% predikční intervaly		t	bodové	95% predikční intervaly	
95	385.250	316.897	453.604	99	297.835	227.802	367.867
96	389.369	320.466	458.272	100	260.154	189.967	330.341
97	372.503	303.105	441.900	101	236.112	165.806	306.419
98	338.994	269.215	408.773	102	232.066	161.596	302.535

3.2.5. Spektrální analýza dat výroby elektrické energie

Již zmíněný postup analýzy časových řad užitý v podkapitolách 3.2.3 a 3.2.4 jsme aplikovali tentokrát na data výroby elektrické energie očištěné od konstantního trendu. Periodogram příslušející těmto datům je zobrazen na obrázku 26. Testové statistiky Fisherova testu a kritické hodnoty jim příslušející jsou uvedeny v tabulce 20. Na hladině významnosti $\alpha = 0.05$ jsme testem určili významné frekvence $\omega_8 = 0.5347$ a $\omega_1 = 0.0668$. Naproti tomu na hladině významnosti $\alpha = 0.01$ nám Fisherův test označil jako významnou pouze frekvenci $\omega_8 = 0.5347$.

Na hladině významnosti 5% jsme vytvořili model skrytých period. Odhad



Obrázek 26: Periodogram dat výroby očištěné od konstantního trendu

Tabulka 20: Fisherův test u dat výroby očištěných od konstantního trendu

m	V	j	$g_F(0.05)$	$g_F(0.01)$	na 5% hladině	na 1% hladině
46	0.6837	8	0.141	0.184	zamítám H_0	zamítám H_0
45	0.1608	1	0.143	0.188	zamítám H_0	nezamítám H_0
44	0.1419	7	0.146	0.192	nezamítám H_0	nezamítám H_0

parametru $\hat{\mu} = -0.2609 \cdot 10^{-12}$ byl kvůli určenému konstantnímu trendu opět zanedbatelný. Přesnost odhadů parametrů $\hat{\alpha}$ a $\hat{\beta}$ je uvedena v tabulce 21. Model skrytých period dat výroby očištěných od konstantního trendu je

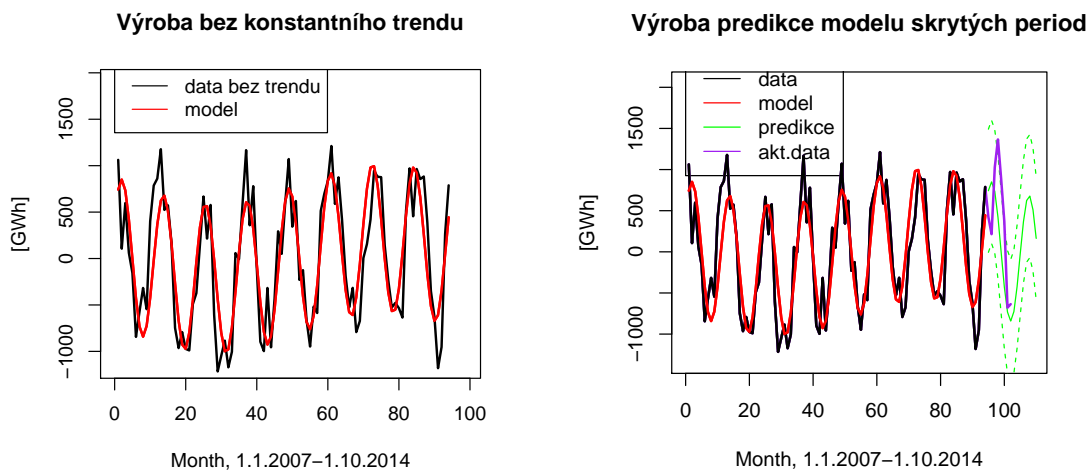
$$\hat{y}_t^{(0)} = 368.3427 \cos(0.5347t) + 713.5605 \sin(0.5347t) + 75.9507 \cos(0.0668t) + 205.4373 \sin(0.0668t).$$

Nalezený model s daty očištěnými od konstantního trendu zobrazuje obrázek 27. Vpravo je model s bodovou a intervalovou predikcí budoucích hodnot a aktuálními daty. Modifikovaný index determinace R_M^2 modelu je 0.7187. Další sledování ukazatelé jsou $RMSE = 357.2606$ a $NRMSE = 0.1383$.

Protože jsme byli na pochybách, zda jsme pomocí Fisherova testu měli určit

Tabulka 21: Přesnost odhadů parametrů modelu skrytých period určených Fisherovým testem dat výroby očištěných od konstantního trendu

Parametr	$\hat{\alpha}_1$	$\hat{\alpha}_8$	$\hat{\beta}_1$	$\hat{\beta}_8$
Odhad	75.9507	368.3427	-205.4373	713.5605
Směrodatná odchylka	53.56	53.56	53.56	53.56



Obrázek 27: Model skrytých period určený Fisherovým testem u dat výroby očištěných od konstantního trendu a jeho predikce

jednu či dvě významné periody, provedli jsme Siegelův test. Na hladině významnosti $\alpha = 0.05$ pro $\lambda = 0.6$ je hodnota testové statistiky

$$T_{0.6} = \sum_{j=1}^{46} (V_j - 0.6g_F)^+ = 0.5991 > 0.0653.$$

Kritická hodnota pro Siegelův test na hladině významnosti $\alpha = 0.05$ byla překročena, a proto jsme zamítli nulovou hypotézu. V součtu zůstala zahrnuta pouze hodnota odpovídající frekvenci $\omega_8 = 0.5347$. Do modelu skrytých period jsme tedy zahrnuli pouze jedinou významnou periodu a vznikl nám model

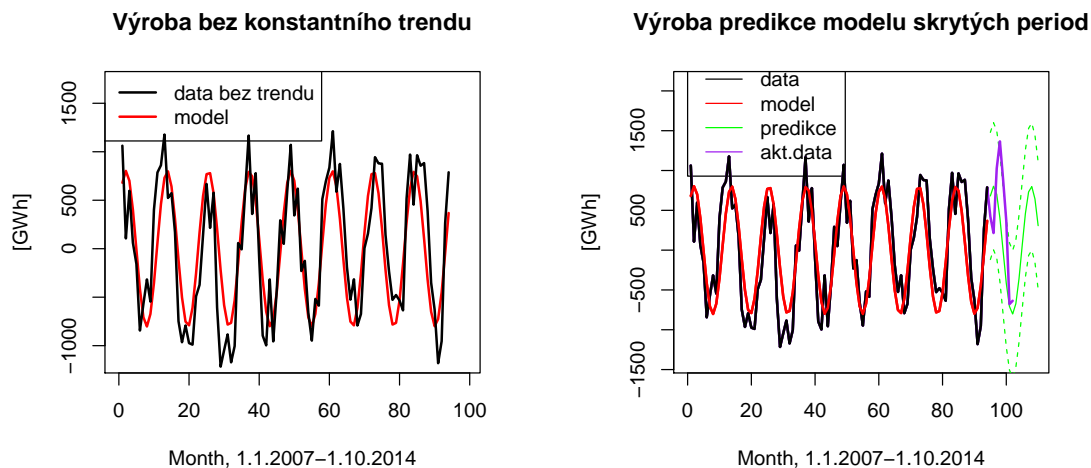
$$\hat{y}_t^{(0)} = 368.3427 \cos(0.5347t) + 713.5605 \sin(0.5347t).$$

Přesnosti odhadů parametrů modelu jsou uvedeny v tabulce 22. Model určený Siegelovým testem je společně s daty výroby očištěnými od konstantního trendu

uveden na obrázku 28. Vpravo je model s predikcí doplněný o aktuální data. Bodové predikce odpovídají aktuálním datům. Aktuální data se v některých částech vyskytují v krajních bodech predikčních intervalů. Charakteristiky modelu jsou $R_M^2 = 0.6731$, $RMSE = 389.3862$ a $NRMSE = 0.1604$.

Tabulka 22: Přesnost odhadů parametrů modelu skrytých period určených Siegelovým testem dat výroby očištěných od konstantního trendu

Parametr	$\hat{\alpha}_8$	$\hat{\beta}_8$
Odhad	368.3427	713.5605
Směrodatná odchylka	57.73	57.73



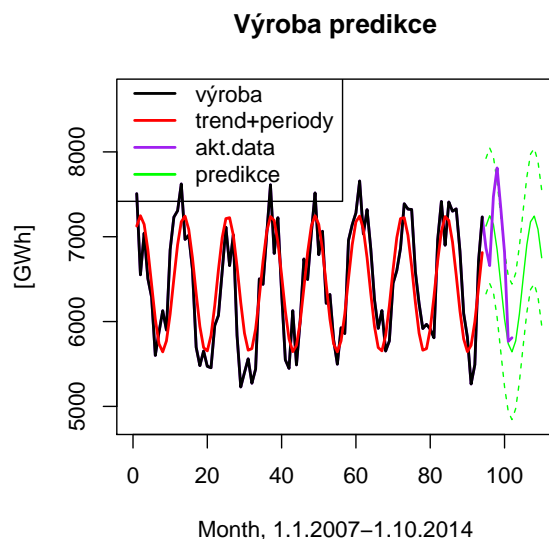
Obrázek 28: Model skrytých period určený Siegelovým testem dat výroby očištěných od konstantního trendu a predikce

Celkový model popisující data výroby je ve tvaru

$$\hat{y}_t = 6444.952 + 368.3427 \cos(0.5347t) + 713.5605 \sin(0.5347t).$$

Výroba elektrické energie zobrazená s konstantním trendem a modelem skrytých period je na obrázku 29. Tento obrázek zobrazuje také bodové a intervalové predikce hodnot určené z nalezeného modelu a je doplněn o aktuální data. Bodové predikce kvalitně zachycují vývoj hodnot aktuálních dat výroby. Aktuální data se vyskytují na mezích predikčních intervalů, ale tuto mez nepřekročují. Model

adekvátně predikuje budoucí hodnoty zkoumané časové řady výroby. Hodnoty bodové predikce a meze intervalových předpovědí jsou uvedeny v tabulce 23. Modifikovaný index determinace modelu je 0.6731. Sledování ukazatelé modelu jsou $RMSE = 389.3862$ a $NRMSE = 0.1604$.



Obrázek 29: Výroba s konstantním trendem, modelem skrytých period určeným Siegelovým testem a predikcí

Tabulka 23: Predikce dat výroby s konstantním trendem a modelem skrytých period

t	bodové	95% predikční intervaly		t	bodové	95% predikční intervaly	
95	7187.577	6438.888	7936.265	99	6442.382	5693.694	7191.070
96	7295.619	6546.931	8044.307	100	6019.495	5270.807	6768.183
97	7179.315	6430.626	7928.003	101	5712.571	4963.883	6461.259
98	6867.185	6118.497	7615.873	102	5603.294	4854.605	6351.982

3.3. Aplikace Boxovy–Jenkinsovy metodologie

Ze začátku jsme chtěli aplikovat Boxovu–Jenkinsovu metodologii přímo na původní ekvidistantní časové řady. Během zpracovávání dat jsme ve všech časových

řadách odhalili sezónnost s periodou o délce 11.75 měsíců, což zhruba odpovídá periodě o délce 1 rok. Sezónnost se v Boxově–Jenkinsově metodologii modeluje stochasticky pomocí *SARIMA* modelů. Při snaze aplikovat *SARIMA* modely jsme ovšem zjistili, že tyto modely kladou větší nároky na počet zpracovávaných dat. *SARIMA* modely se totiž tvoří rozčleněním původní časové řady na řady pro konkrétní období. V našem případě bychom vytvořili nejdříve model pro lednová pozorování, následně model pro únorová pozorování a tak dále. Vzniklo by nám tedy dvanáct oddělených časových řad a každou bychom modelovali zvlášť. Tyto modely by byly pro jednotlivé měsíce podobné. Následně bychom vytvořili na základě korelované náhodné složky těchto modelů model jediný, který by zachytil všechny časové řady současně. Při rozčlenění této časové řady o délce 94 pozorování, by nám pro modelování například lednové složky zbývalo pouze 8 hodnot, což pro hledání vhodného modelu není dostačující. Pro modelování každé části časové řady bychom potřebovali data o délce přibližně 50 pozorování. Pro měsíční data je tedy tento postup vhodný až pro data o délce například 600 a více. Naše časové řady neobsahují dostatečný počet členů pro aplikaci *SARIMA* modelů, a proto jsme od této metody museli upustit.

Další možností jak postupovat by mohlo být využití diferencování. Pokud bychom tyto časové řady zkoušeli pouze diferencovat, museli bychom pro eliminaci sezónní složky využít vysokého řádu diferencování. Nedoporučuje se však diferencovat časové řady více jak dvakrát, protože při vhodném určení řádu diferencování klesá rozptyl dané časové řady. Pokud ovšem budeme diferencovat příliš mnohokrát, rozptyl začne neúměrně narůstat. Každé diferencování také vede ke zkreslení původních dat a čím více bychom diferencovali, tím bychom více zkreslili výsledky provedené analýzy. Výsledky by neměly vhodnou vypovídací schopnost. Proto ani diferencování není vhodné aplikovat na naše analyzované časové řady.

Vzhledem k výše uvedenému jsme k analýze původních ekvidistantních časových řad přistoupili následovně. Data jsme očistili od trendové a periodické složky, které jsme již našli v podkapitolách 3.2.1 až 3.2.5. Ta část časové řady,

která nebyla zachycena předchozími metodami, byla předmětem dalšího zkoumání. Takto vzniklé dílčí části časové řady budou označeny $y_t^{(NS)}$. Tuto analýzu jsme provedli v programu *R*, kde jsme použili již naprogramované funkce pro vytvoření *ARMA* modelů. Pokud dále v textu uvedeme sousloví „modely *ARMA*“ nebo „*ARMA* modely“, tak tím bude myšlena obecně skupina modelů $AR(p)$, $MA(q)$, $ARMA(p, q)$ a jejich konkrétní verze pro hodnoty parametrů p a q .

Při výstavbě modelů *ARMA* jsme museli nejdříve projít fázi identifikace modelu. V této části jsme data vykreslili pomocí funkce *plot()* a ověřili si jejich stacionaritu. Časové řady jsou již centrované díky provedenému očištění. Následně jsme určili hodnoty autokorelační a parciální autokorelační funkce. V programu *R* jsme hodnoty těchto funkcí získali užitím následujících příkazů

```
# vykreslí hodnoty autokorelační funkce
rk=acf(epst, type=c("correlation"), demean=TRUE, plot=TRUE)
rk$acf # vypíše hodnoty autokorelace
# vykreslí parciální autokorelační funkci
pautokor=pacf(epst, plot=TRUE)
pautokor$acf # vypíše hodnoty parciální autokorelace
```

Ve fázi identifikace modelu jsme zkoumali průběh autokorelační funkce a parciální autokorelační funkce. Hledali jsme jejich identifikační body, pomocí nichž jsme se snažili odhalit vhodný model. Následoval odhad parametrů modelu, k čemuž jsme využili v programu *R* funkci *arima()*:

```
model=arima(epst, order=c(1,0,1))
```

Správnost nalezeného modelu jsme ověřili ve fázi verifikace, kde jsme využili funkce *tsdiag()* a *forecast()*:

```
tsdiag(model)
plot(forecast(model,h=12))
```

Užitím příkazu *tsdiag()* jsme získali tři grafy. První graf zobrazuje standardizovaná rezidua, druhý výsledky testování Metodou odhadnutých reziduí.

Poslední z těchto grafů zobrazuje výsledky Ljungovy–Boxovy statistiky. Příkaz `plot(forecast(model, h = 12))` nám zobrazí bodovou a intervalovou predikci nalezeného modelu. Konkrétní aplikace tohoto příkazu je zobrazena např. na obrázku 34.

3.3.1. Boxova–Jenkinsova metodologie pro data dovozu

Boxovu–Jenkinsovu metodologii jsme v této části aplikovali na data očištěná od nalezených modelů popisujících trendovou a periodickou složku časových řad. V kapitole 3.2.3 jsme pro data dovozu určili aplikací spektrální analýzy dva vhodné modely. Prvním byl model s konstantním trendem a modelem skrytých period s dvěma významnými frekvencemi ω_1 a ω_8 . Data pro další zpracování jsme získali očištěním dat dovozu od tohoto modelu, což zapíšeme vztahem

$$y_t - \hat{y}_t = y_t - 795.80 - 162.0834 \cos(0.0668t) - 95.5208 \sin(0.0668t) + \\ - 75.6021 \cos(0.5347t) + 137.6084 \sin(0.5347t).$$

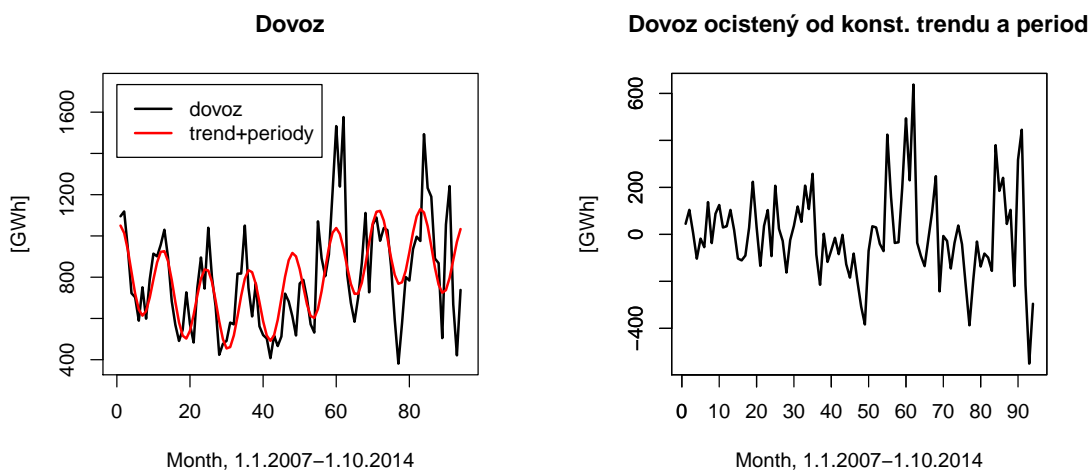
Nově vzniklou časovou řadu si pro jednodušší identifikaci označíme MODEL 1. V modelech si neuvádíme odhadnutou hodnotu $\hat{\mu}$ z důvodu její zanedbatelnosti. Druhým modelem, který vhodně popisuje data dovozu, je model s polynomickým trendem 3. řádu a modelem skrytých period s významnými frekvencemi ω_3 a ω_8 . Časovou řadu pro aplikaci Boxovy–Jenkinsovy metodologie jsme získali z následujícího vztahu

$$y_t - \hat{y}_t = y_t - 1030.524 + 33.177t - 0.822t^2 + 0.005t^3 - 69.4041 \cos(0.2005t) + \\ + 63.9603 \sin(0.2005t) - 163.4846 \cos(0.5347t) - 86.7345 \sin(0.5347t).$$

Takto vzniklou časovou řadu označíme MODEL 2. Boxovu–Jenkinsovu metodologii jsme aplikovali na MODEL 1 i na MODEL 2. Tuto podkapitolu si tedy rozdělíme na dvě části. V první části popíšeme hledání nejvhodnějšího modelu *ARMA* pro časovou řadu MODELU 1, ve druhé části pak stejný postup zopakujeme pro časovou řadu MODELU 2. Nakonec pro názornost nalezené modely porovnáme.

Boxova–Jenkinsova metodologie pro MODEL 1

Časovou řadu dovozu jsme aditivně dekomponovali zvlášť trendovou a periodickou složku a zvlášť náhodnou složku. V levé části obrázku 30 jsou zobrazena původní data dovozu, druhá křivka na grafu vlevo pak zobrazuje již popsanou část časové řady, tj. konstantní trend s modelem skrytých period se dvěma významnými frekvencemi ω_1 a ω_8 (na obrázku označeny „trend+periody“). Na grafu vpravo je zobrazena pouze křivka náhodné složky (dále jen MODEL 1), pro kterou jsme hledali vhodný model *ARMA*.



Obrázek 30: Aditivní dekompozice časové řady dovozu pro konstantní trend s modelem skrytých period a náhodnou složkou

Časovou řadu MODELU 1 bychom mohli označit za stacionární. Úvahu jsme si ověřili vykreslením autokorelační funkce MODELU 1, která je zobrazena na obrázku 31 vlevo. Protože hodnoty ACF jsou rychle klesající, můžeme časovou řadu MODELU 1 označit za stacionární. Testováním hypotézy

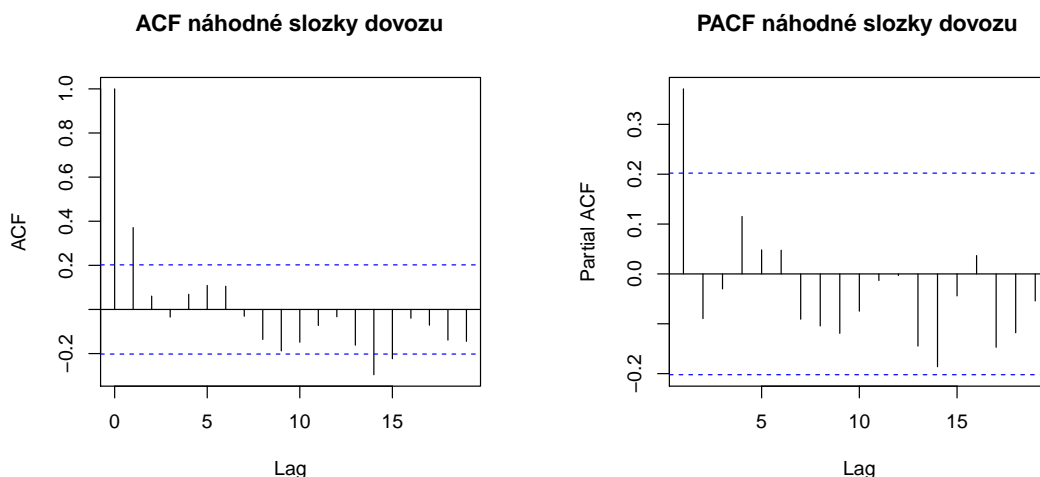
$$H_0 : r_k = 0, \quad k > k_0$$

jsme hledali identifikační bod k_0 , od kterého jsou odhadnuté hodnoty autokorelační funkce MODELU 1 dostatečně blízko nule. Hodnoty odhadnuté autokorelační funkce jsme porovnávali s hodnotou $2\sigma(r_k)$ určenou pomocí vzorce (12).

Hodnoty, které stanovenou mez překročily, mají indexy $k = 1, 14$ a 15 . Překročení meze na pozici 14 a 15 může být způsobeno chybou měření nebo chybou odhadu. Zaměřili jsme se proto pouze na hodnotu odhadnuté autokorelační funkce s indexem $k = 1$, tzn. že identifikační bod k_0 je roven 1. V pravé části obrázku 31 jsou zobrazeny odhadnuté hodnoty parciální autokorelační funkce MODELU 1. V případě hledání identifikačního bodu pro parciální autokorelační funkci MODELU 1 jsme testovali hypotézu

$$H_0 : r_{kk} = 0, \quad k > k_0.$$

K nalezení identifikačního bodu jsme využili hodnotu $2\sigma(r_{kk})$ získanou ze vztahu (13). Stanovenou mez překročuje pouze první hodnota parciální autokorelační funkce, identifikační bod PACF je $k_0 = 1$.



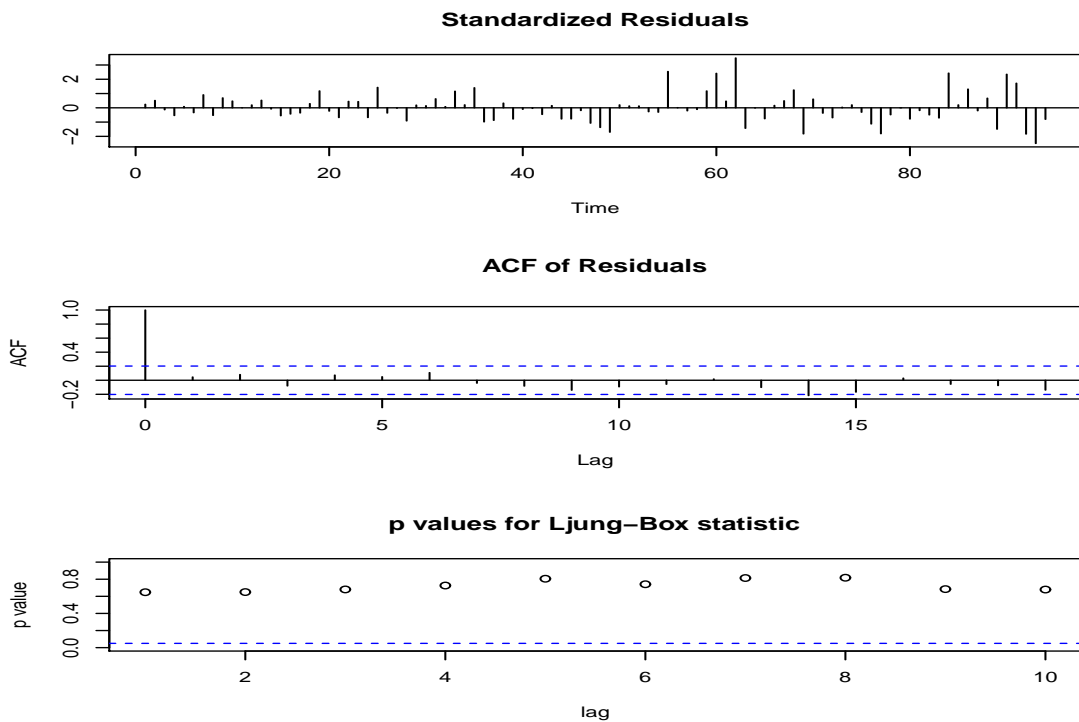
Obrázek 31: ACF a PACF MODELU 1

Identifikační body směřují k určení modelu $ARMA(1, 1)$. Hodnoty autokorelační funkce a parciálně autokorelační funkce MODELU 1 jsou pouze odhady, proto jsme ověřili i vhodnost modelů $AR(1)$ a $MA(1)$. Využili jsme k tomu v programu *R* příkaz `arima(data, order = c(1, 0, 1))`. Odhady parametrů modelu $ARMA(1, 1)$ jsou uvedeny v tabulce 24. V tabulce jsou i hodnoty směrodatných odchylek odhadnutých parametrů a hodnota Akaikeho informačního kritéria. Hodnoty směrodatných odchylek odhadů parametrů modelu $ARMA(1, 1)$

jsou velmi vysoké a to nám značí malou přesnost odhadů parametrů. Zaměřili jsme se proto raději na modely $AR(1)$ a $MA(1)$. Odhad parametru a směrodatné odchyly je pro model $AR(1)$ uveden ve třetím sloupci tabulky 24. Hodnoty příslušející modelu $MA(1)$ jsou uvedeny v posledním sloupci této tabulky. Nejmenší hodnota AIC svědčí pro model $AR(1)$.

Tabulka 24: Model $ARMA(1, 1)$, $AR(1)$ a $MA(1)$ MODELU 1

Model	$ARMA(1, 1)$		$AR(1)$	$MA(1)$
	$\hat{\varphi}_1$	$\hat{\theta}_1$	$\hat{\varphi}_1$	$\hat{\theta}_1$
Odhady parametrů	0.2462	0.1524	0.3774	0.3535
Směrodatné odchyly	0.2313	0.2276	0.0962	0.0845
AIC	1244.3		1242.71	1243.34



Obrázek 32: Ověření modelu $MA(1)$ MODELU 1

Nalezené modely $MA(1)$ i $AR(1)$ jsme dále ověřili z hlediska jejich adekvátnosti. K ověření jsme užívali Metodu odhadnutých reziduí a Ljungovu-Boxovu

statistiku. U Metody odhadnutých reziduí jsme pro model $MA(1)$ porovnávali odhadnuté hodnoty autokorelační funkce odhadnutých reziduí ($r_k(\hat{\varepsilon})$) s dvojnásobkem jejich směrodatné odchylky. Pokud pro některou hodnotu $r_k(\hat{\varepsilon})$ platí

$$|r_k(\hat{\varepsilon})| > 2\sigma\{r_k(\hat{\varepsilon})\},$$

pak bychom měli zvážit, zda je model $MA(1)$ vhodný a ověřit jej ještě jiným testem. Výsledné zhodnocení metody jsme získali užitím příkazu *tsdiag()* v programu *R*. Na obrázku 32 je v prostřední části zobrazena autokorelační funkce odhadnutých reziduí modelu $MA(1)$. K překročení dvojnásobku směrodatné odchylky odhadů došlo na pozici 14. Jelikož jsme si nebyli jisti vhodností zvoleného modelu $MA(1)$, ověřili jsme jeho vhodnost ještě Portmanteau testem. Test hodnotí souhrnné korelační chování odhadnutých reziduí. Hodnotu jeho testovací statistiky jsme vypočítali pomocí vztahu

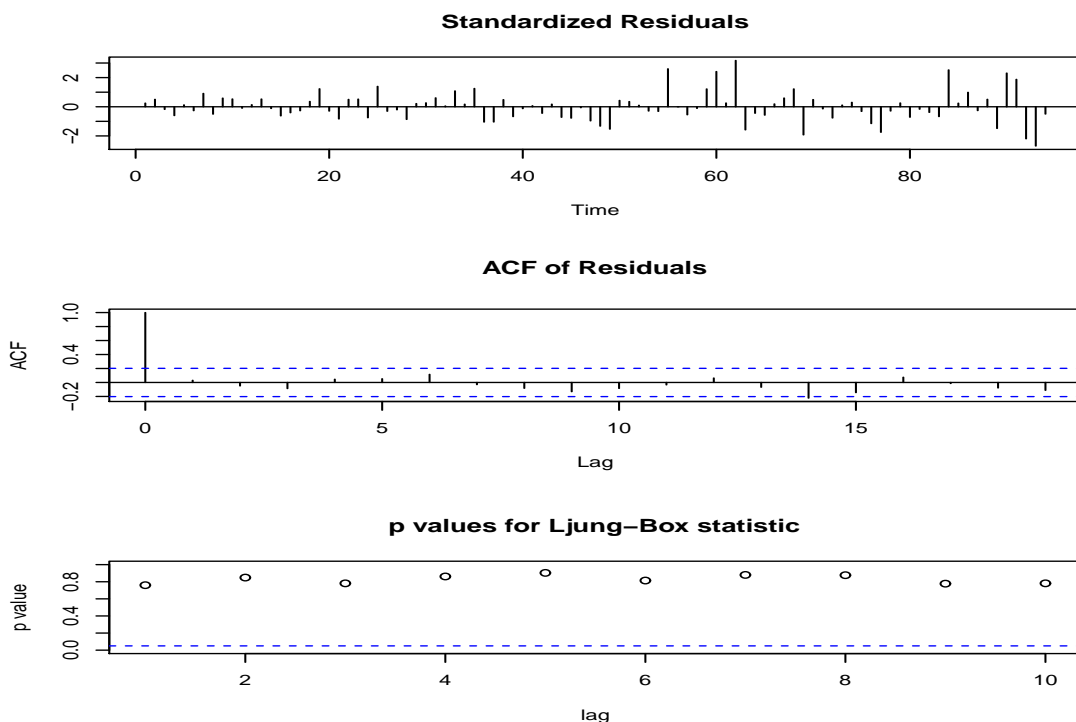
$$Q = n \sum_{k=1}^K r_k^2(\hat{\varepsilon}) = 94 \cdot [(0.046)^2 + (0.082)^2 + \dots + (-0.093)^2] = 6.8017,$$

kde $K = \sqrt{n} = \sqrt{94} = 9.6954 \doteq 10$. Hodnotu Q jsme porovnávali s kritickou hodnotou $\chi_{10-1}^2(0.05) = 16.919$. Platí $Q < \chi_9^2(0.05)$, tj. není překročena kritická hodnota na hladině významnosti α a určený model $MA(1)$ je adekvátní. Pomocí programu *R* jsme dále ověřovali vhodnost modelu $MA(1)$ pomocí Ljungovy–Boxovy statistiky, kterou značíme Q^* a je modifikací Portmanteau testu. Platí pro ni vztah

$$Q^* = n(n+2) \sum_{k=1}^K \frac{r_k^2(\hat{\varepsilon})}{n-k}.$$

Model bychom označili za nevhodný, pokud by hodnota Q^* na hladině významnosti α překročila kritickou hodnotu $\chi_{K-p-q}^2(\alpha)$, kde $K = \sqrt{n} \doteq 10$. Výsledky testování jsme opět získali užitím funkce *tsdiag()*, jejíž výsledky jsou pro model $MA(1)$ zobrazeny na obrázku 32 v posledním grafu. Funkce zobrazuje přerušovanou čarou hladinu významnosti α , pod kterou nesmí klesnout hodnoty p-value

statistiky Q^* . Pokud jsou hodnoty p-value menší než hladina významnosti α , model označíme za nevhodný. Zobrazené hodnoty p-value jsou výrazně nad hladinou významnosti α , a proto jsme potvrdili, že model $MA(1)$ je vhodný.



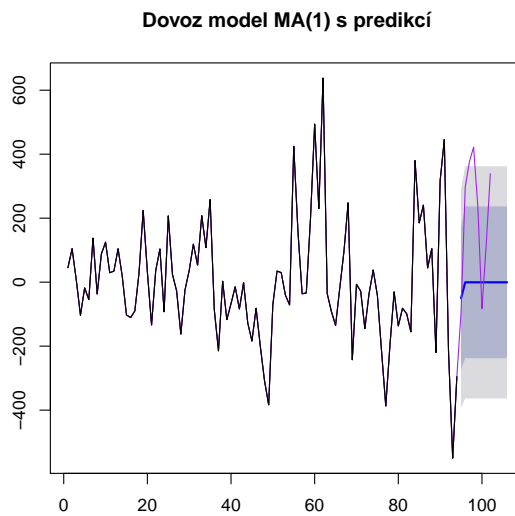
Obrázek 33: Ověření modelu $AR(1)$ MODELU 1

Ověření vhodnosti modelu $AR(1)$ je zobrazeno na obrázku 33. Metoda odhadnutých reziduí opět prokázala překročení meze na pozici 14. Ověřením kvality modelu pomocí Ljungovy-Boxovy statistiky žádná z hodnot p-value neklesla pod hladinu významnosti α , ani se jí nepřiblížila. Model $AR(1)$ jsme také označili za vhodný. Při zkoumání významnosti překročení mezí u Metody odhadnutých reziduí jsme zjistili, že k většímu překročení mezí došlo u modelu $AR(1)$. Jako nejvhodnější model pro popis očištěných dat dovozu jsme zvolili model $MA(1)$ ve tvaru

$$y_t^{(NS)} = \varepsilon_t + 0.3535\varepsilon_{t-1}.$$

Pro model $MA(1)$ jsme zobrazili bodové a intervalové predikce na obrázku 34,

který je doplněn o očištěná aktuální data (zobrazená fialovou barvou). Očištěná aktuální data jsme získali odečtením bodové predikce uvedené v tabulce 10 od aktuálních dat. Intervalové odhady modelu $MA(1)$, až na drobné překročení mezí, očištěná aktuální data zachytily, proto jsme mohli predikci považovat za adekvátní pro zachycení očištěných aktuálních dat. Hodnoty bodových predikcí i meze intervalových předpovědí jsou uvedeny v tabulce 25.



Obrázek 34: Predikce z modelu $MA(1)$ MODELU 1

Tabulka 25: Předpověď z modelu $MA(1)$ MODELU 1

t	bodové	80% predikční intervaly		95% predikční intervaly	
95	-50.1813	-273.8145	173.4518	-392.1987	291.8361
96	-0.4738	-237.6702	236.7227	-363.2344	362.2869
97	-0.4738	-237.6702	236.7227	-363.2344	362.2869
98	-0.4738	-237.6702	236.7227	-363.2344	362.2869

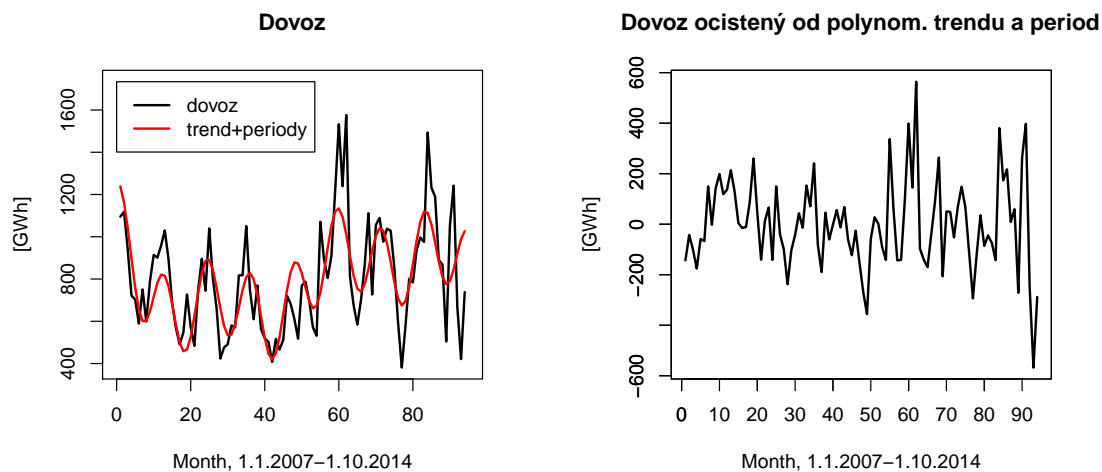
Konečný model popisující data dovozu se skládá z konstantního trendu, modelu skrytých period s významnými frekvencemi ω_1 a ω_8 a modelu $MA(1)$. Tento

model je ve tvaru

$$y_t = 795.80 + 75.60 \cos(0.53t) + 162.08 \cos(0.07t) - 137.61 \sin(0.53t) + \\ + 95.52 \sin(0.07t) + \varepsilon_t + 0.3535\varepsilon_{t-1}.$$

Boxova–Jenkinsova metodologie pro MODEL 2

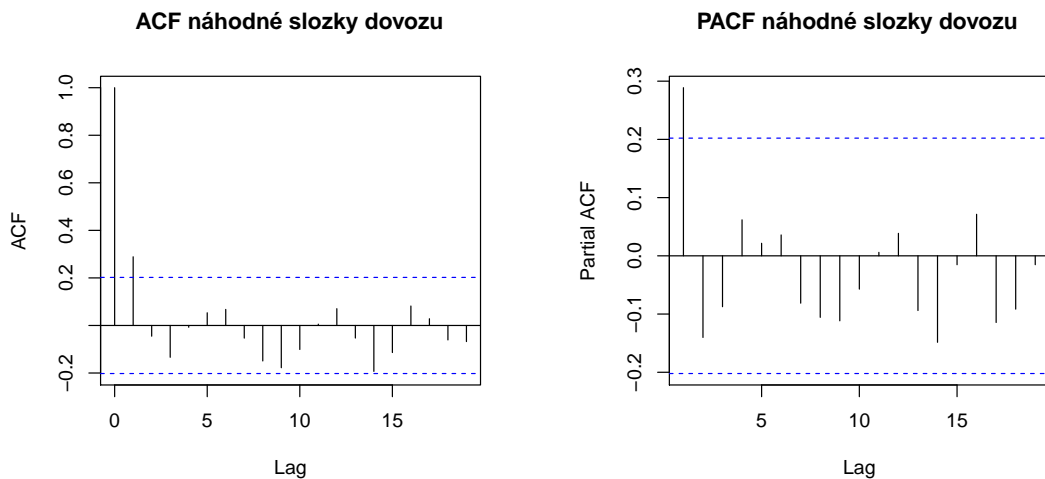
Všechny fáze výstavby *ARMA* modelů si zopakujeme také pro časovou řadu MODELU 2, tj. pro data dovozu elektrické energie očištěná od polynomického trendu 3. řádu a modelu skrytých period s významnými frekvencemi ω_3 a ω_8 . Toto očištění dat dovozu je zobrazeno na obrázku 35. V levé části obrázku 35 se nachází graf původních dat dovozu, přičemž červená křivka zobrazuje polynomický trend třetího řádu s modelem skrytých period. Data, která jsme později modelovali pomocí *ARMA* modelů, jsou zobrazena na grafu vpravo. Jde o část časové řady dat dovozu, která není zachycena předchozími metodami a označovali jsme ji MODEL 2.



Obrázek 35: Aditivní dekompozice časové řady dovozu pro polynomický trend 3. řádu s modelem skrytých period a náhodnou složkou

Časová řada MODELU 2 se zdá být stacionární. Pro ověření stacionarity jsme si vykreslili průběh autokorelační funkce MODELU 2, který je zobrazen na obrázku 36 vlevo. Hodnoty ACF jsou rychle klesající, a proto jsme časovou

řadu MODELU 2 označili za stacionární. Hranici $2\sigma(r_k)$ překročila pouze první hodnota odhadnuté autokorelační funkce, identifikační bod ACF je tedy $k_0 = 1$. Vpravo na obrázku 36 je zobrazen průběh odhadnuté parciální autokorelační funkce MODELU 2. Stanovenou mez překračuje pouze první hodnota PACF. Identifikační bod odhadnuté parciální autokorelační funkce je $k_0 = 1$.

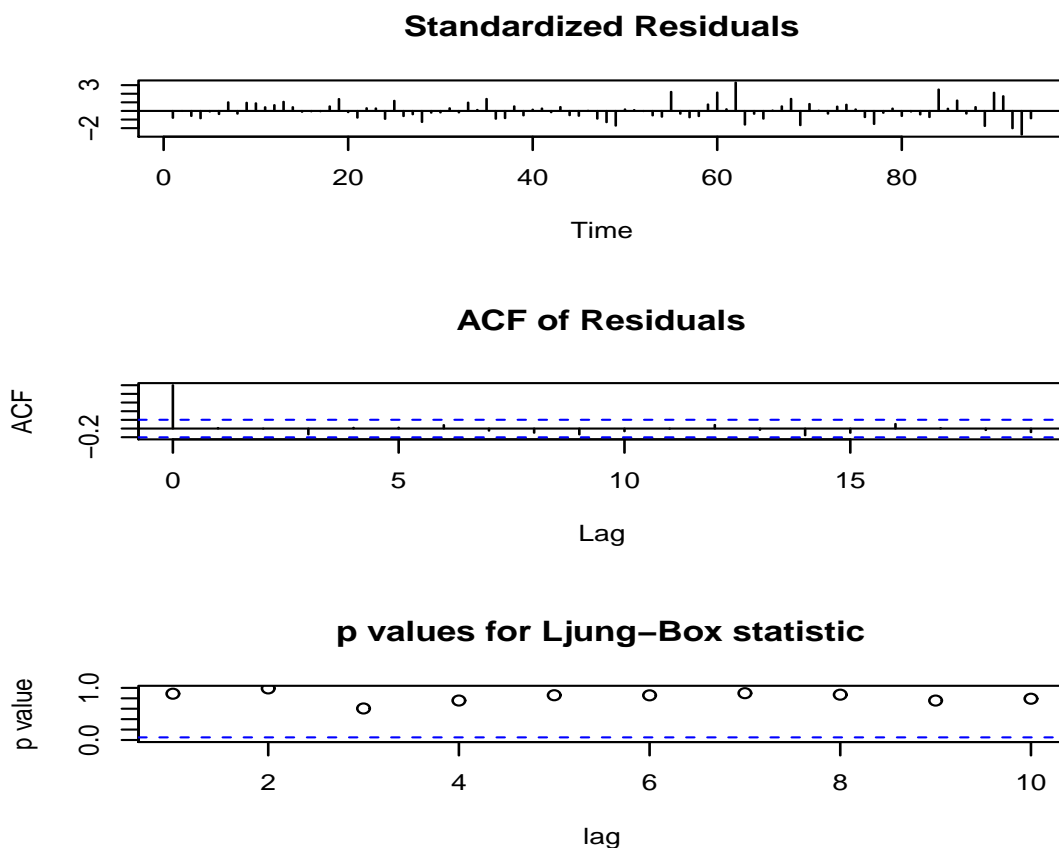


Obrázek 36: ACF a PACF MODELU 2

Nalezeným identifikačním bodům funkcí ACF a PACF odpovídají modely $ARMA(1,1)$, $AR(1)$ a $MA(1)$. Použitím příkazu `arima()` jsme získali odhady parametrů, směrodatné odchytky i hodnoty AIC , vše je zapsáno v tabulce 26. Směrodatné odchytky odhadů parametrů modelu $ARMA(1,1)$ jsou vyšší než samotné odhady parametrů, proto model $ARMA(1,1)$ není vhodný. Dále jsme se rozhodovali mezi modely $MA(1)$ a $AR(1)$. Díky nižší hodnotě Akaikeho informačního kritéria jsme zvolili jako vhodný model $MA(1)$.

Tabulka 26: Model $ARMA(1,1)$, $AR(1)$, $MA(1)$ MODELU 2

Model	$ARMA(1,1)$		$AR(1)$	$MA(1)$
	$\hat{\varphi}_1$	$\hat{\theta}_1$	$\hat{\varphi}_1$	$\hat{\theta}_1$
Odhady parametrů	0.0848	0.2384	0.2964	0.3090
Směrodatné odchytky	0.2539	0.2369	0.0997	0.0909
AIC	1237.76		1236.66	1235.86



Obrázek 37: Ověření modelu $MA(1)$ MODELU 2

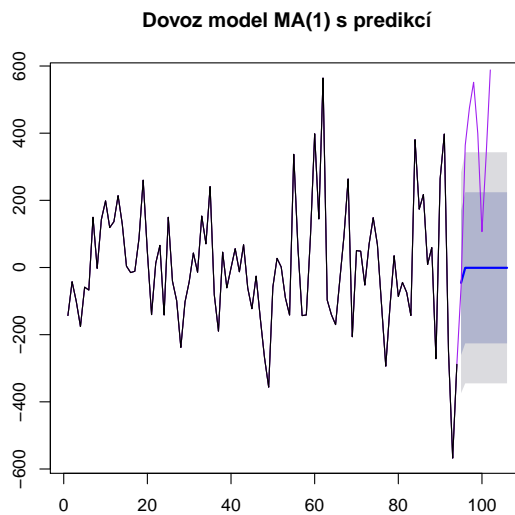
Výsledky z verifikace modelu $MA(1)$ jsou zobrazeny na obrázku 37. Prostřední graf zobrazuje výsledky Metody odhadnutých reziduí, žádná z hodnot $r_k(\hat{\varepsilon})$ nepřekročila dvojnásobek jejich směrodatné odchylky (zobrazen přerušovanou čarou). To potvrzuje, že model $MA(1)$ je vhodný pro zachycení časové řady MODELU 2. Hodnoty p-value zobrazené v posledním grafu na obrázku 37 neklešají pod hladinu významnosti α , a proto i Ljungova-Boxova statistika označila model $MA(1)$ za adekvátní pro modelování časové řady MODELU 2.

Model $MA(1)$ je tvaru

$$y_t^{(NS)} = \varepsilon_t + 0.3090\varepsilon_{t-1}.$$

Pro model $MA(1)$ jsme vytvořili bodovou a intervalovou predikci, která je zob-

razená na obrázku 38. Hodnoty očištěných aktuálních dat (zobrazeny fialovou barvou) jsme získali očištěním aktuálních dat dovozu od bodové predikce uvedené v tabulce 14. Očištěná aktuální data se nacházejí mimo predikční intervaly, to může být způsobeno právě očištěním aktuálních dat od bodové predikce. Jelikož hodnotíme kvalitu predikce modelu $MA(1)$ pro zachycení budoucího vývoje dat MODELU 2, museli jsme v tomto případě predikci modelu $MA(1)$ označit za nedostatečně kvalitní. Hodnoty bodových a meze intervalových předpovědí jsou uvedeny v tabulce 27.



Obrázek 38: Predikce z modelu $MA(1)$ MODELU 2

Tabulka 27: Předpověď z modelu $MA(1)$ MODELU 2

t	bodové	80% predikční intervaly		95% predikční intervaly	
95	-45.6249	-260.5778	169.3279	-374.3670	283.1171
96	-0.9158	-225.8951	224.0635	-344.9920	343.1603
97	-0.9158	-225.8951	224.0635	-344.9920	343.1603
98	-0.9158	-225.8951	224.0635	-344.9920	343.1603

Celkový model zachycující data dovozu je ve tvaru

$$y_t = 1030.52 - 33.18t + 0.82t^2 - 0.005t^3 + 69.40 \cos(0.20t) + 163.48 \cos(0.53t) + \\ - 63.96 \sin(0.20t) + 86.73 \sin(0.53t) + \varepsilon_t + 0.31\varepsilon_{t-1}.$$

Srovnání MODELU 1 a MODELU 2

Očištěním dat dovozu od modelů nalezených v podkapitole 3.2.3 nám vznikly jiné časové řady, které odpovídaly náhodným složkám jednotlivých modelů pro původní časové řady. Tyto jsme si nazvali MODEL 1 a MODEL 2. Na obě řady jsme aplikovali Boxovu–Jenkinsovu metodologii a hledali jsme vhodné *ARMA* modely. Dosažené výsledky si zde porovnáme.

V případě dat z MODELU 1 překročily tři hodnoty odhadnuté autokorelační funkce stanovenou mez. U dat dovozu z MODELU 2 došlo u funkce ACF jen k jednomu překročení mezí. Identifikačním bodem funkcí ACF jsme v obou případech určili $k_0 = 1$. Pro odhadnuté parciální autokorelační funkce jsme shodně určili identifikační bod $k_0 = 1$. V obou případech jsme směřovali k určení modelu *ARMA*(1, 1) jako nejvhodnějšího pro zachycení dat obou modelů. U obou modelů byly velké směrodatné odchylky odhadů parametrů. Pro zachycení časových řad MODELU 1 i MODELU 2 jsme ve fázi verifikace dali nakonec přednost modelům *MA*(1).

Tabulka 28: Porovnání nalezených modelů náhodných složek dat dovozu

Data dovozu	MODEL 1	MODEL 2
Model	<i>MA</i> (1)	<i>MA</i> (1)
Odhady parametrů	$\hat{\theta}_1 = 0.3535$	$\hat{\theta}_1 = 0.3090$
Směrodatné odchylky	0.0845	0.0909
<i>AIC</i>	1243.34	1235.86

Odhady parametrů modelů *MA*(1) pro MODEL 1 i MODEL 2 jsou uvedeny v tabulce 28. Nižší hodnota *AIC* svědčí pro model *MA*(1) zachycující časovou řadu MODELU 2. Při verifikaci došlo u modelu *MA*(1) časové řady MODELU 1 u Metody odhadnutých reziduí k překročení stanovené meze na pozici 14. Z těchto

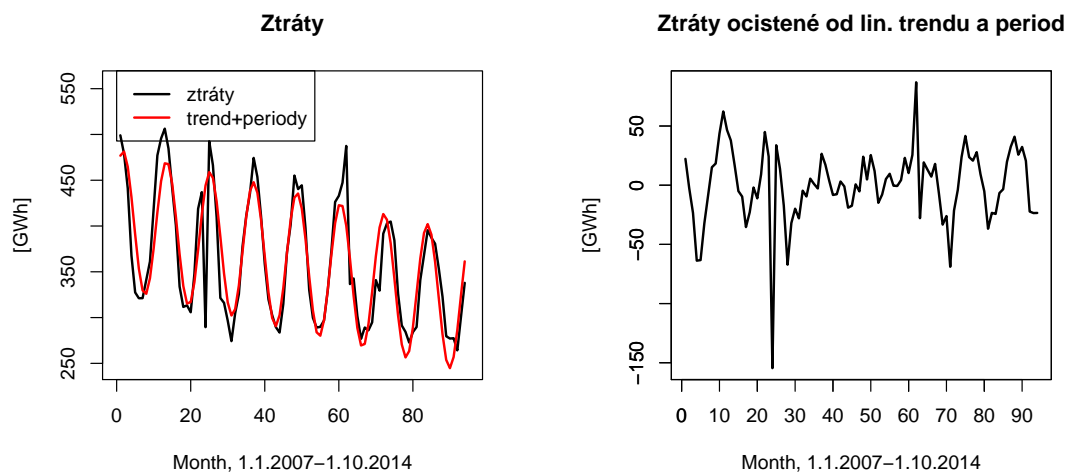
důvodů můžeme hodnotit pro popis náhodné složky dat dovozu jako vhodnější model $MA(1)$ MODELU 2. Při srovnávání očištěných aktuálních hodnot s vytvořenou predikcí nám očištěné aktuální hodnoty lépe zachytily predikční intervaly modelu $MA(1)$ příslušející MODELU 1. Tento výsledek byl bohužel ovlivněn očištěním aktuálních hodnot od bodových predikcí uvedených v tabulkách 10 a 14. Pokud bychom si ale museli jeden z výsledných modelů vybrat, tak bychom pro popis dat dovozu nakonec volili $MA(1)$ pro MODEL 1.

3.3.2. Boxova–Jenkinsova metodologie pro data ztrát

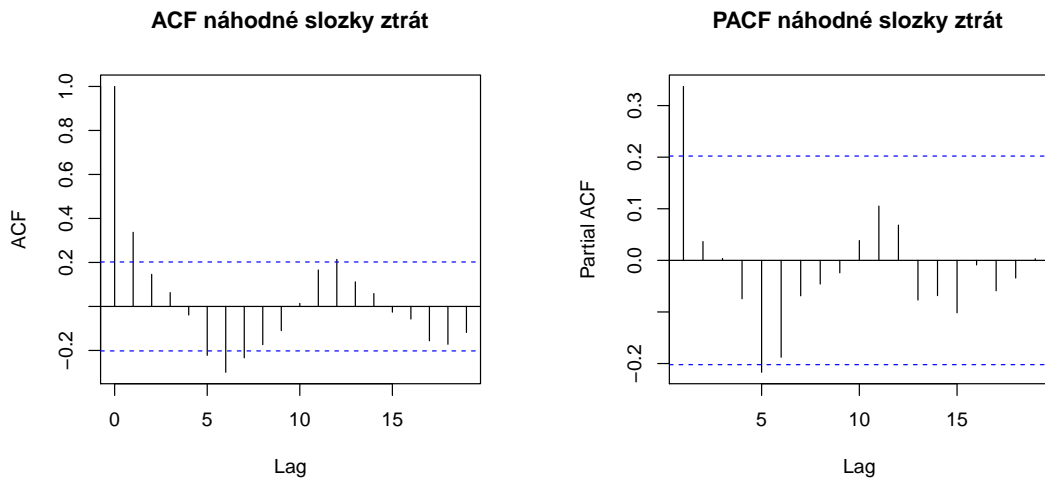
Data ztrát elektrické energie v sítích jsme očistili od lineárního trendu a jedné významné periody nalezené pomocí Siegelova testu, což zapíšeme

$$y_t - \hat{y}_t = y_t - 408.07 + 0.97t - 44.84 \cos(0.53t) - 61.12 \sin(0.53t).$$

Takto vzniklou časovou řadu si označíme MODEL 3. Očištění časové řady dat ztrát jsme si zobrazili na obrázku 39. V levé části obrázku 39 jsou zobrazena původní data ztrát a označení „trend+periody“ používáme pro nalezený lineární trend a model skrytých period. Druhý graf na obrázku 39 zobrazuje náhodnou složku dat ztrát označenou jako MODEL 3.



Obrázek 39: Aditivní dekompozice dat ztrát pro lineární trend s modelem skrytých period určeným Siegelovým testem a náhodnou složkou



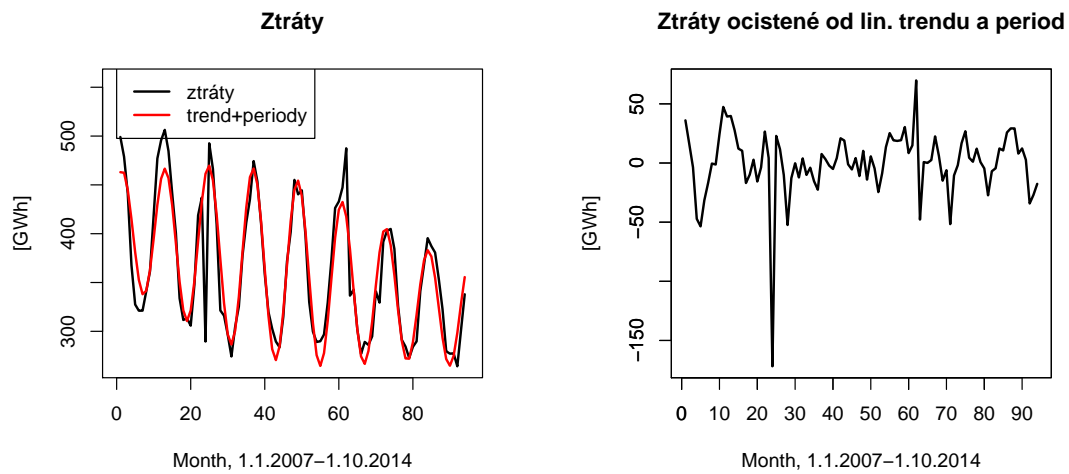
Obrázek 40: ACF a PACF MODELU 3

Stacionaritu časové řady MODELU 3 ověříme zobrazením hodnot odhadnuté autokorelační funkce na obrázku 40. Z funkce ACF je patrné, že časová řada z MODELU 3 ještě obsahuje periodickou složku. Hodnoty ACF mají tvar sinusovky s pomalu klesající amplitudou. Z těchto důvodů jsme opět původní časovou řadu dat ztrát očistili od lineárního trendu a modelu skrytých period s významnými frekvencemi ω_7 a ω_8 , které jsme nyní určili pomocí Fisherova testu. Nově nám tedy vznikl jiný model, který je tentokrát ve tvaru

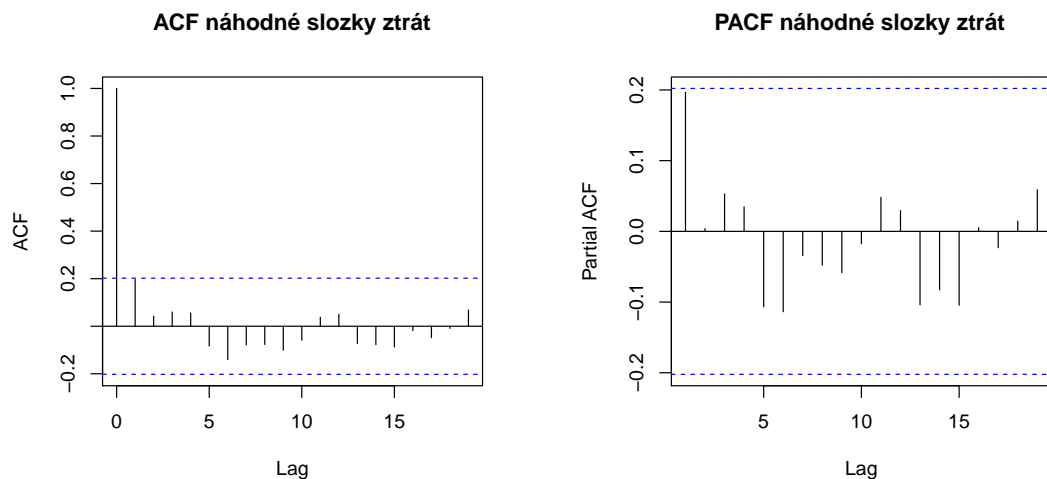
$$y_t - \hat{y}_t = y_t - 408.07 + 0.97t + 5.81 \cos(0.47t) + 19.14 \sin(0.47t) + \\ - 44.84 \cos(0.53t) - 61.12 \sin(0.53t).$$

Označíme ho MODEL 4. Znovu si zobrazíme původní ekvidistantní časovou řadu dat ztrát na obrázku 41. Červenou barvou je zobrazen lineární trend s modelem skrytých period nalezeným s využitím Fisherova testu. Vpravo je zobrazena časová řada MODELU 4, na kterou aplikujeme Boxovu–Jenkinsovu metodologii.

Stacionaritu časové řady MODELU 4 jsme ověřili zobrazením funkce ACF na obrázku 42 vlevo. Rychle klesající odhadnuté hodnoty autokorelační funkce nasvědčují stacionaritě MODELU 4. Odhadnutá autokorelační funkce má identifikační bod $k_0 = 1$. Odhadnutá parciální autokorelační funkce, která je zobrazena na obrázku 42 vpravo, má identifikační bod $k_0 = 0$, protože žádná z hodnot PACF



Obrázek 41: Aditivní dekompozice dat ztrát pro lineární trend s modelem skrytých period určeným Fisherovým testem a náhodnou složkou



Obrázek 42: ACF a PACF MODELU 4

nepřekročila zobrazenou mez.

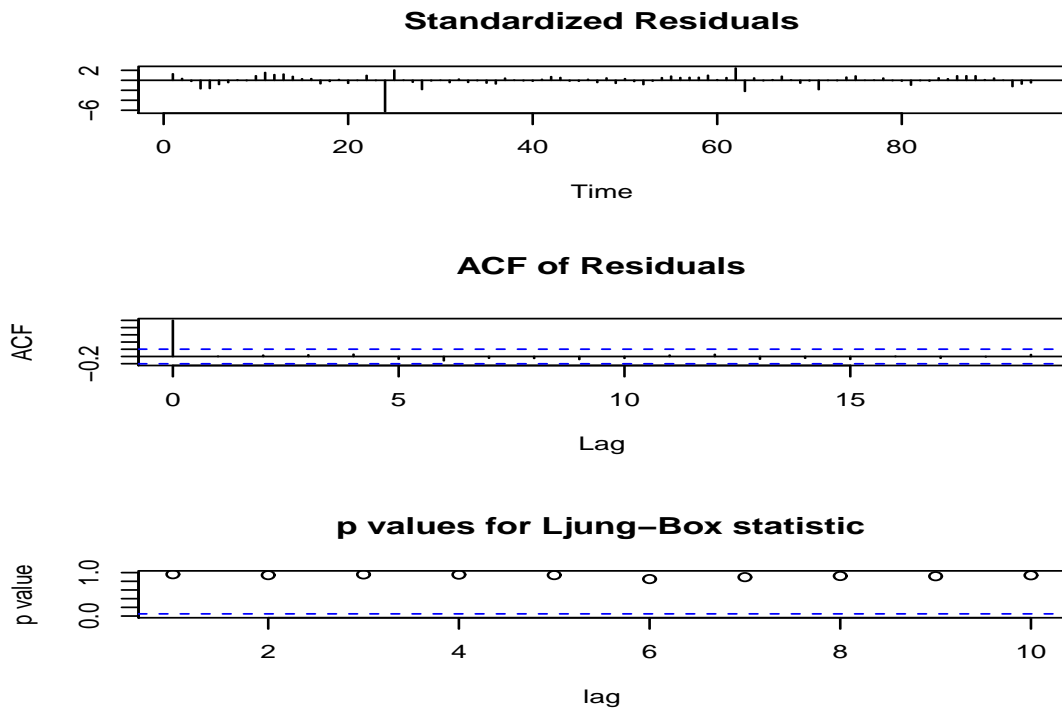
MODEL 4 nejlépe zachycuje model $MA(1)$. Odhad parametru modelu $MA(1)$ a jeho směrodatná odchylka jsou uvedeny v tabulce 29. Vhodnost použití modelu $MA(1)$ jsme ověřili Metodou odhadnutých reziduí a Ljungovou–Boxovou statistikou. Výstupy obou ověření jsou zobrazeny na obrázku 43. Prostřední graf obrázku 43 zobrazuje, že hodnoty $r_k(\hat{\varepsilon})$ modelu $MA(1)$ nepřekročí a ani se nepřiblíží

Tabulka 29: Model $MA(1)$ MODELU 4

Model	$MA(1)$
Odhad parametru	$\hat{\theta}_1 = 0.1935$
Směrodatná odchylka	0.0989
AIC	895.47

dvojnásobku jejich směrodatné odchylky (zobrazen přerušovanou čarou). Po ověření vhodnosti modelu $MA(1)$ Metodou odhadnutých reziduí jsme jej mohli označili za vhodný. Hodnoty p-value Ljungovy–Boxovy statistiky neklesají pod hladinu významnosti α , což zobrazuje třetí graf na obrázku 43. Tato statistika potvrzuje adekvátnost modelu $MA(1)$. Model $MA(1)$ zachycující data z MODELU 4 zapíšeme jako

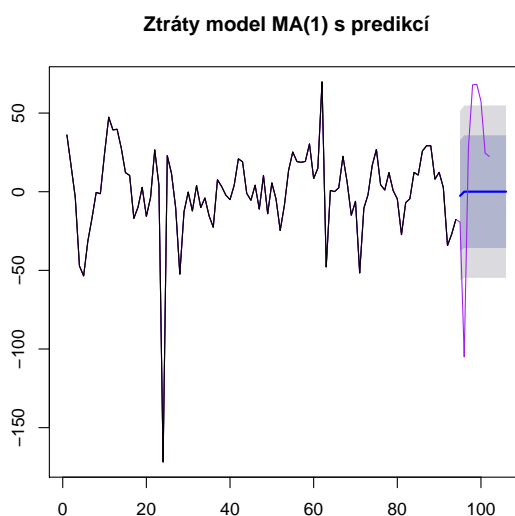
$$y_t^{(NS)} = \varepsilon_t + 0.1935\varepsilon_{t-1}.$$



Obrázek 43: Ověření modelu $MA(1)$ MODELU 4

Na obrázku 44 jsou zobrazeny bodové a intervalové predikce dat z modelu

$MA(1)$. Očištěná aktuální data ztrát (zobrazena fialovou barvou) jsme získali odečtením bodové predikce uvedené v tabulce 19 od aktuálních hodnot dat ztrát. Meze intervalových předpovědí byly očištěnými aktuálními daty překročeny. Predikce hodnot z modelu $MA(1)$ nezachytila budoucí vývoj očištěných aktuálních dat dostatečně přesně. Predikované hodnoty z modelu $MA(1)$ jsou uvedeny v tabulce 30.



Obrázek 44: Predikce z modelu $MA(1)$ MODELU 4

Tabulka 30: Predikce z modelu $MA(1)$ MODELU 4

t	bodové	80% predikční intervaly		95% predikční intervaly	
95	-2.6155	-37.7837	32.5527	-56.4006	51.1696
96	0.0397	-35.7806	35.8600	-54.7427	54.8222
97	0.0397	-35.7806	35.8600	-54.7427	54.8222
98	0.0397	-35.7806	35.8600	-54.7427	54.8222

Náhodnou složku dat ztrát popsanou MODELEM 3, jsme tedy nemodelovali pomocí $ARMA$ modelů. Určili jsme vhodný model pouze pro časovou řadu popsanou MODELEM 4. Určený model $MA(1)$ pro MODEL 4 zahrneme do celkového modelu. Celkový model popisující data ztrát elektrické energie v sítích je

tedy ve tvaru

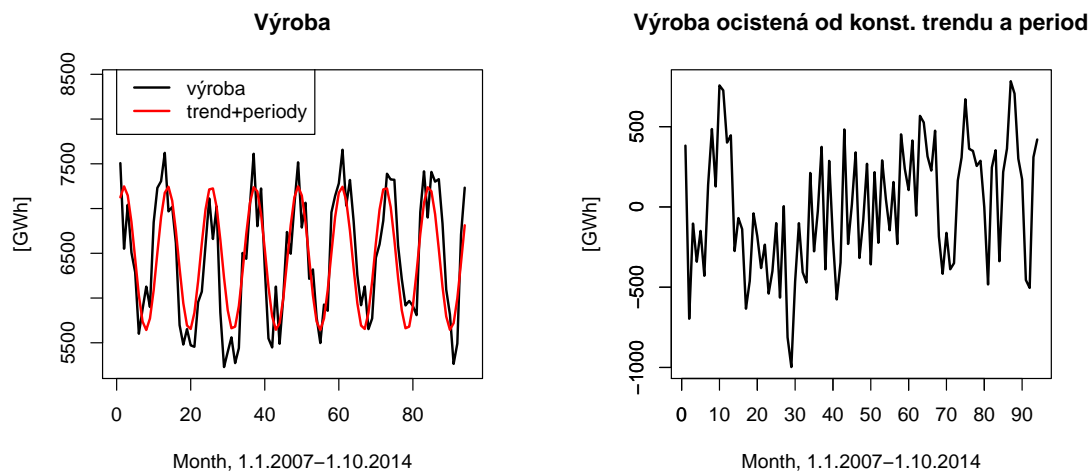
$$y_t = 408.07 - 0.97t + 44.84 \cos(0.53t) + 61.12 \sin(0.53t) - 5.81 \cos(0.47t) - 19.14 \sin(0.47t) + \varepsilon_t + 0.194\varepsilon_{t-1}.$$

3.3.3. Boxova–Jenkinsova metodologie pro data výroby

Data výroby elektrické energie jsme očistili od konstantního trendu a významných period určených podle Siegelova testu, což zapíšeme jako

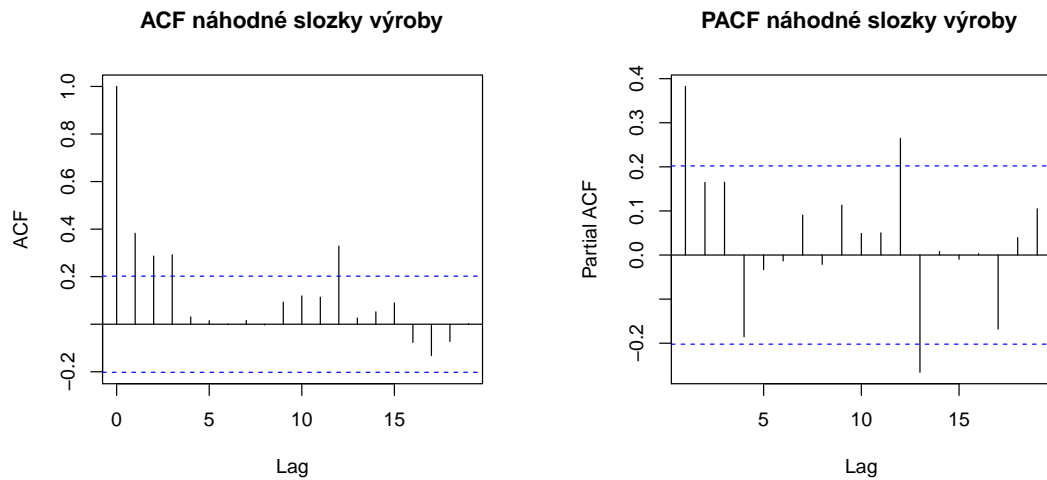
$$y_t - \hat{y}_t = y_t - 6444.95 - 368.34 \cos(0.53t) - 713.56 \sin(0.53t).$$

Časovou řadu, která vznikla tímto očištěním, si označíme MODEL 5. Vlevo na obrázku 45 jsou původní data výroby proložená červenou křivkou konstantního trendu a modelem skrytých period s významnou frekvencí ω_8 . Očištěním dat výroby nám vznikla časová řada MODELU 5, která je zobrazená vpravo na obrázku 45. Pokusili jsme se ji zachytit pomocí vhodného modelu *ARMA*.



Obrázek 45: Aditivní dekompozice časové řady výroby pro konstantní trend s modelem skrytých period a náhodnou složkou

Časovou řadu MODELU 5 můžeme označit jako stacionární. Na obrázku 46 vlevo je zobrazen průběh odhadnuté autokorelační funkce MODELU 5. Hodnoty



Obrázek 46: ACF a PACF MODELU 5

ACF, které překročily mez, mají indexy 1, 2, 3 a 12. Pokud bychom jako identifikační bod ACF určili hodnotu 12, došlo by ke zbytečnému zesložiténí modelu. K překročení meze mohlo dojít chybou během očištění původních dat výroby. Jako identifikační bod odhadnuté autokorelační funkce jsme tedy zvolili $k_0 = 3$. Identifikační bod pro odhadnutou parciální autokorelační funkci, která je zobrazená na obrázku 46 vpravo, jsme určili $k_0 = 1$. K překročení mezí odhadnuté PACF v bodech $k = 12, 13$ mohlo dojít chybou během očištění dat nebo chybou při získávání dat.

Tabulka 31: Model $ARMA(1, 3)$ MODELU 5

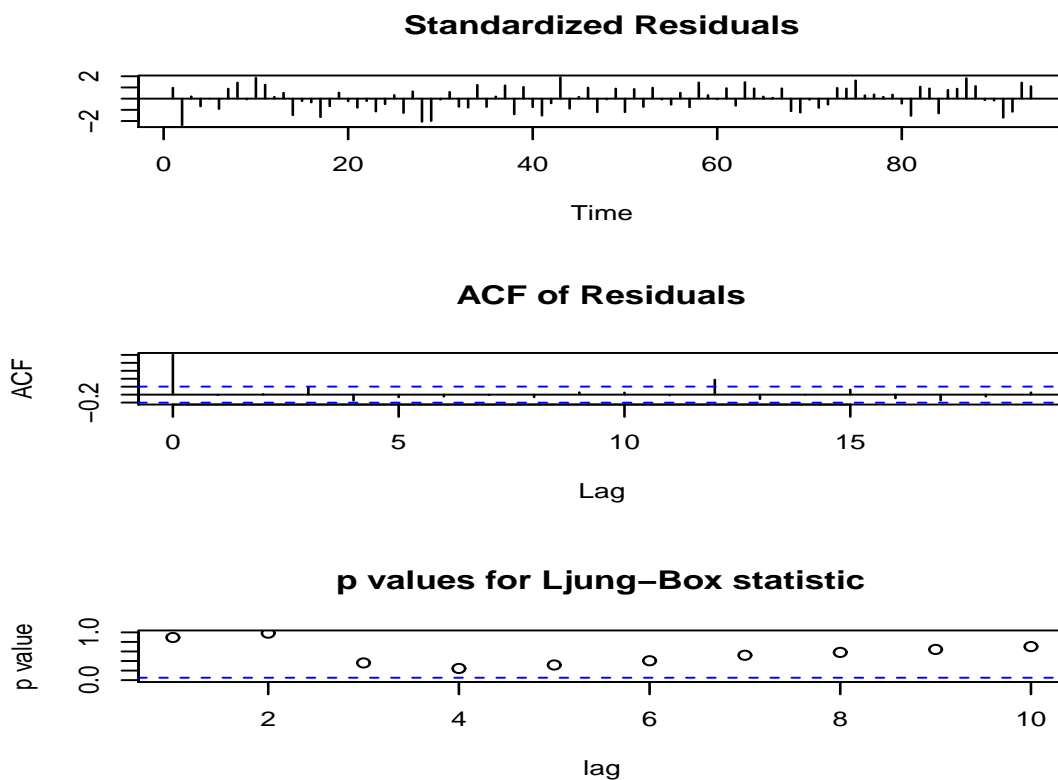
Model	$ARMA(1, 3)$			
	$\hat{\varphi}_1$	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\theta}_3$
Odhady parametrů	0.0129	0.3171	0.2620	0.3895
Směrodatné odchytky	0.2408	0.2184	0.1272	0.1026
AIC	1374.67			

Nejvhodnějším modelem pro zachycení hodnot řady MODELU 5 by mohl být model $ARMA(1, 3)$. Odhady parametrů modelu $ARMA(1, 3)$ jsou i s jejich směrodatnými odchylkami uvedeny v tabulce 31. Hodnoty směrodatných odchylek modelu $ARMA(1, 3)$ jsou příliš vysoké, model $ARMA(1, 3)$ je navíc pro popis

MODELU 5 příliš složitý. Proto jsme hledali vhodnější model.

Tabulka 32: Model $ARMA(1, 1)$ pro data z MODELU 5

Model	$ARMA(1, 1)$	
	$\hat{\varphi}_1$	$\hat{\theta}_1$
Odhady parametrů	0.7050	-0.3808
Směrodatné odchytky	0.1415	0.1800
AIC	1378.09	



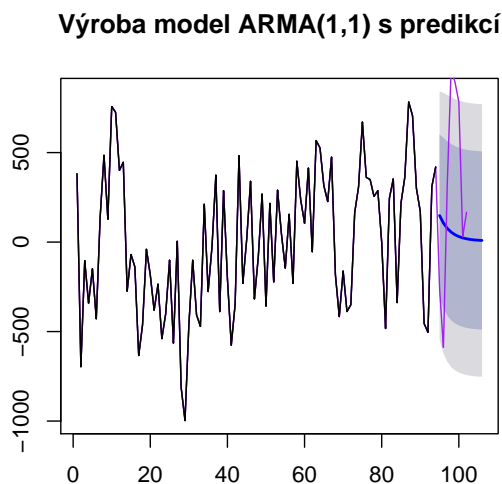
Obrázek 47: Ověření modelu $ARMA(1, 1)$ MODELU 5

Po ověřování několika dalších modelů jsme jako nejvhodnější model pro zachycení dat MODELU 5 určili model $ARMA(1, 1)$. Odhad parametrů modelu $ARMA(1, 1)$ uvádí tabulka 32. Výsledky verifikace modelu $ARMA(1, 1)$ zobrazuje obrázek 47. V prostředním grafu tohoto obrázku je zobrazeno na pozici 12

překročení dvojnásobku směrodatné odchyly odhadnutých hodnot autokorelační funkce modelu $ARMA(1, 1)$. Metodou odhadnutých reziduí jsme tedy nepotvrdili adekvátnost modelu $ARMA(1, 1)$. Hodnoty p-values Ljungovy–Boxovy statistiky, zobrazené na obrázku 47 na spodním grafu, neklesají pod hladinu významnosti α . Model tedy můžeme, díky výsledkům Ljungovy–Boxovy statistiky, označit za vhodný.

Model $ARMA(1, 1)$ zachycující data z MODELU 5 je ve tvaru

$$y_t^{(NS)} = 0.7050y_{t-1} + \varepsilon_t - 0.3808\varepsilon_{t-1}.$$



Obrázek 48: Predikce z modelu $ARMA(1, 1)$ MODELU 5

Na obrázku 48 jsou zobrazeny hodnoty bodové predikce a meze predikčních intervalů budoucích hodnot z modelu $ARMA(1, 1)$. Očištěná aktuální data (zobrazena fialovou barvou) jsme získali očištěním aktuálních dat výroby od bodové predikce uvedené v tabulce 23. Predikční intervaly jsou mírně překročeny, model $ARMA(1, 1)$ ale dostatečně kvalitně predikuje budoucí vývoj očištěných aktuálních dat. V tabulce 33 jsou uvedeny hodnoty bodových a intervalových předpovědí modelu $ARMA(1, 1)$.

Tabulka 33: Predikce z modelu $ARMA(1, 1)$ MODELU 5

t	bodové	80% predikční intervaly		95% predikční intervaly	
95	148.4362	-304.4609	601.3334	-544.2102	841.0826
96	106.4070	-369.6985	582.5124	-621.7335	834.5474
97	76.7779	-410.4504	564.0062	-668.3735	821.9292
98	55.8904	-436.7722	548.5531	-697.5721	809.3530
99	41.1656	-454.1757	536.5068	-716.3935	798.7246
100	30.7851	-465.8820	527.4521	-728.8016	790.3717

Konečný model vhodně charakterizující data výroby je tvaru

$$y_t = 6444.95 + 368.34 \cos(0.53t) + 713.56 \sin(0.53t) + 0.7050y_{t-1} + \varepsilon_t - 0.3808\varepsilon_{t-1}.$$

3.4. Shrnutí výsledků

Výsledky dosažené v předchozích částech v této podkapitole znovu uvedeme a zhodnotíme. Aplikací spektrální analýzy jsme data dovozu zachytili pomocí tří modelů, jejichž charakteristiky jsou uvedeny v tabulce 34. Označení F jsme použili pro model skrytých period určený pomocí Fisherova testu a S označuje určení modelu skrytých period pomocí Siegelova testu. V prvním řádku tabulky 34 jsou uvedeny charakteristiky pro model I. s konstantním trendem a modelem skrytých period s významnými frekvencemi ω_1 a ω_8 . Ve druhém řádku jsou ukazatelé pro model II. s polynomickým trendem třetího řádu a model skrytých period s významnou frekvencí ω_8 . Poslední řádek obsahuje charakteristiky modelu III. s polynomickým trendem třetího řádu a modelu skrytých period s významnými frekvencemi ω_3 a ω_8 . Největší hodnoty modifikovaného indexu determinace dosahují data dovozu modelovaná pomocí modelu III. Ukazatelé $RMSE$ a $NRMSE$ modelu III. mají nejmenší hodnotu, proto je model III. nejlepší pro zachycení dat dovozu. Intervalové predikce modelu III. jsou v porovnání s modelem I. více neurčité. Lepších výsledků jsme při predikování budoucích hodnot dosáhli u modelu I. Toto tvrzení také potvrzují nejmenší hodnoty MAE , $MAPE$ a přesnější predikční intervaly zachycující budoucí vývoj dat dovozu u modelu I.

Tabulka 34: Charakteristiky kvality modelů pro data dovozu

Model (data dovozu)	R_M^2	$RMSE$	$NRMSE$	MAE	$MAPE$
I. konst. trend +F	0.4358	187.6949	0.1571	245.6175	23.1553
II. polynom. trend 3. ř.+F	0.4248	188.4595	0.1577	339.2525	32.2885
III. polynom. trend 3. ř.+S	0.4852	176.2472	0.1475	359.5187	35.1421

Pro data ztrát jsme obdobné srovnání uvedli v tabulce 35. Zde má větší modifikovaný index determinace model IV. s lineárním trendem a modelem skrytých period se dvěma významnými frekvencemi ω_7 a ω_8 . Jako vhodnější model jsme ovšem určili ten druhý, tedy model V. s lineárním trendem a modelem skrytých period s významnou frekvencí ω_8 . Model V. jsme určili Siegelovým testem, který při složené periodicitě dosahuje kvalitnějších výsledků než Fisherův test. Z modelu V. jsme vytvořili bodovou a intervalovou predikci hodnot. Predikční intervaly až na drobné odchylky aktuální data ztrát kvalitně zachytily.

Tabulka 35: Charakteristiky kvality modelů pro data ztrát

Model (data ztrát)	R_M^2	$RMSE$	$NRMSE$	MAE	$MAPE$
IV. lineární trend+F	0.8183	27.9825	0.1156	49.1074	15.5088
V. lineární trend+S	0.7769	31.3551	0.1295	50.3171	16.4351

Poslední zkoumanou časovou řadou byla data výroby. Ukazatelé kvality modelu VI. s konstantním trendem a modelem skrytých period s významnými frekvencemi ω_1 a ω_8 jsou uvedeny v prvním řádku tabulky 36. Druhý řádek tabulky 36 obsahuje charakteristiky modelu VII. s konstantním trendem a modelem skrytých period s významnou frekvencí ω_8 . Hodnota R_M^2 je vyšší pro model VI. a hodnoty $RMSE$ i $NRMSE$ jsou také lepší pro tento model, ale opět jsme dali přednost modelu VII. pro zachycení dat výroby a při vytváření predikcí. Bodové predikce z modelu VII. totiž dobře zachycují vývoj aktuálních dat, což potvrzují i nižší hodnoty MAE a $MAPE$. Predikční intervaly modelu VII. jsou relativně úzké, ovšem aktuální data výroby do predikčních intervalů náleží, a proto model dobře vystihnul budoucího chování dat výroby.

Když jsme měli popsány u jednotlivých časových řad trendové a periodické

Tabulka 36: Charakteristiky kvality modelů pro data výroby

Model (data výroby)	R_M^2	$RMSE$	$NRMSE$	MAE	$MAPE$
VI. konstantní trend+F	0.7187	357.2606	0.1471	516.9209	7.3086
VII. konstantní trend+S	0.6731	389.3862	0.1604	500.9715	7.0422

složky, následovala analýza nepopsaných částí časových řad, které jsme označili jako náhodnou složku časové řady. U dat dovozu jsme očištěním dat od modelů I. a II. získali časové řady MODELU 1 a MODELU 2, na které jsme aplikovali Boxovu–Jenkinsovu metodologii. Na základě postupů uvedených v podkapitole 3.3.1 jsme pro popis časových řad MODELU 1 a MODELU 2 jako nejvhodnější určili model $MA(1)$. Charakteristiky hodnotící kvalitu obou modelů jsou uvedeny v tabulce 37. Hodnota Akaikeho informačního kritéria označila jako lepší model $MA(1)$ popisující MODEL 2. Dle střední absolutní chyby je lepší model $MA(1)$ popisující MODEL 1, který také dosahuje lepších výsledků při intervalové predikci. Hodnocení podle MAE a $MAPE$ je bohužel ovlivněno očištěním aktuálních dat od bodové predikce modelů I. a III.

Tabulka 37: Porovnání nalezených modelů náhodných složek dat dovozu

Data dovozu	MODEL 1	MODEL 2
Model	$MA(1)$	$MA(1)$
Odhady parametrů	$\hat{\theta}_1 = 0.3535$	$\hat{\theta}_1 = 0.3090$
Směrodatné odchylky	0.0845	0.0909
AIC	1243.3400	1235.8600
MAE	239.6409	354.6169
$MAPE$	93.4265	90.5665

Boxovu–Jenkinsovu metodologii jsme chtěli aplikovat na očištěná data ztrát z MODELU 3. V této časové řadě zůstala ale zachycena periodická složka, a proto jsme Boxovu–Jenkinsovu metodologii aplikovali na časovou řadu MODELU 4. V tomto případě je označení MODEL 3 a 4 převzat z podkapitoly 3.3.2. Pomocí verifikace jsme ověřili, že nejvhodnější model popisující časovou řadu MODELU 4 je model $MA(1)$. Charakteristiky kvality modelu $MA(1)$ jsou uvedeny v tabulce 38. Meze intervalových predikcí z modelu $MA(1)$ byly očištěnými aktu-

álními hodnotami překročeny. Bodové a intervalové predikce hodnot nejsou tedy dostatečně kvalitní, hodnocení této predikce je ale ovlivněno očištěním aktuálních dat ztrát od bodové predikce modelu IV.

Tabulka 38: Model $MA(1)$ MODELU 4

Model	$MA(1)$
Odhad parametru	$\hat{\theta}_1 = 0.1935$
Směrodatná odchylka	0.0989
AIC	895.4700
MAE	48.7557
$MAPE$	98.2255

Poslední zkoumanou časovou řadou byla náhodná složka dat výroby, kterou jsme označili MODEL 5, viz podkapitola 3.3.3. Ve fázi verifikace jsme určili jako nejvhodnější model popisující časovou řadu MODELU 5 model $ARMA(1, 1)$. Odhady parametrů a ukazatele kvality modelu $ARMA(1, 1)$ jsou uvedeny v tabulce 39. Intervalová predikce modelu byla očištěnými aktuálními daty mírně překročena, ale vývoj dat zachytila dostatečně dobře.

Tabulka 39: Model $ARMA(1, 1)$ MODELU 5

Model	$ARMA(1, 1)$	
	$\hat{\varphi}_1$	$\hat{\theta}_1$
Odhad parametru	0.7050	-0.3808
Směrodatná odchylka	0.1415	0.1800
AIC	1378.0900	
MAE	502.0276	
$MAPE$	93.3495	

Závěr

Diplomová práce vznikla za účelem analýzy dat Energetického regulačního úřadu. První kapitola, která mimo jiné odkazuje na související legislativu regulace energetického trhu, obsahuje popis struktury ERÚ a historického vývoje úřadu.

Ve druhé kapitole je uvedena teorie potřebná k vysvětlení metod použitých k analýze dat. K datům jsme přistupovali jako k časovým řadám, které jsme aditivně dekomponovali na složku trendovou, periodickou a náhodnou. Trendovou složku jsme popsali analyticky některými trendovými křivkami. Periodickou složku jsme zkoumali pomocí spektrální analýzy. Vytvořili jsme modely skrytých period pomocí Fisherova testu a jeho modifikace. Z nalezených modelů jsme také predikovali budoucí hodnoty časových řad. Náhodnou složku jsme analyzovali pomocí Boxovy–Jenkinsovy metodologie. Představili jsme si proto obecný lineární proces i jeho speciální případy, proces klouzavých součtů (MA) a autoregresní proces (AR). Popsány byly také smíšený proces $ARMA$, modely $ARIMA$ umožňující stochastický popis trendové složky a modely $SARIMA$ analyzující stochasticky sezónní složku časové řady. Dále byly vysvětleny fáze výstavby $ARMA$ modelů a teoretický postup pro získání předpovědí budoucích hodnot časové řady v $ARMA$ modelech.

Třetí kapitola již obsahuje samotnou aplikaci matematického aparátu na konkrétní data získaná z internetových stránek Energetického regulačního úřadu. Z několika možných časových řad jsme si vybrali takové tři, jejichž hodnoty si i běžný čtenář dokáže představit. Jde o data dovozu elektrické energie do České republiky, ztrát elektrické energie v distribučních sítích a výrobu elektrické energie na území České republiky. Jedná se o měsíční data za období od ledna 2007 do října 2014. V době zpracování diplomové práce byla data postupně doplněna o hodnoty od listopadu 2014 do června 2015. Nově zveřejněné hodnoty jsme však do analýzy nezahrnuli. Naopak jsme je využili pro srovnání kvality předpovědí z námi nalezených modelů. Veškerá analýza byla provedena v programu R . Zdrojové kódy algoritmů naleznete na přiloženém datovém nosiči.

Cílem diplomové práce bylo analyzovat data, sestavit vhodný model a ná-

sledně vytvořit z tohoto modelu predikci budoucích hodnot. Užili jsme k tomu Spektrální analýzu a *ARMA* modely. Obě metody jsme chtěli aplikovat na data odděleně, což se nám povedlo jen částečně. Důvodem byla sezónnost dat a dále jejich nedostatečný počet pro analýzu dat pomocí *SARIMA* modelů. Proto jsme k analýze přistoupili jiným způsobem. Nejdříve jsme data očistili od trendu a periodické složky a *ARMA* modely jsme aplikovali na očištěná data. Výsledky dosažené během analýzy dat jsou shrnuty a zhodnoceny v kapitole 3.4.

Závěrem je třeba konstatovat, že výsledky dosažené Spektrální analýzou jsou dostatečně kvalitní. Bodové predikce se jen mírně odchyľují od reálných dat a intervalové predikce zachytily aktuální data uvnitř intervalů. Výsledky dosažené aplikací modelů *ARMA* bohužel nemají dostatečně dobrou vypovídající schopnost, protože jsme data museli pro jejich užití nejprve očistit. Z tohoto důvodu došlo k výraznému zkreslení výsledků. Tato data by bylo vhodné modelovat *ARMA* modely až po uplynutí několika let.

Seznam obrázků

1	Průběh ρ_{kk} procesu MA(q) (zdroj: [5, str. 114])	34
2	Určení transformační konstanty λ (zdroj: [5, str. 145])	43
3	Dovoz energie	61
4	Dovoz proložený trendy	64
5	Charakteristiky parametru konstantního trendu dovozu	64
6	Charakteristiky parametrů polynomického trendu 3. řádu dovozu	65
7	Dovoz očištěn od konstantního trendu a jeho predikce	66
8	Dovoz proložen polynomickým trendem 3. řádu a jeho predikce . .	66
9	Ztráty proložené trendy	67
10	Charakteristiky parametrů lineárního trendu dat ztrát	68
11	Ztráty proložené lineárním trendem a jejich predikce	68
12	Výroba proložená trendy	69
13	Významnost parametru konstantního trendu výroby	70
14	Konstantní trend a predikce výroby	70
15	Periodogram dat dovozu očištěných od konstantního trendu . . .	72
16	Model skrytých period a jeho predikce u dat dovozu očištěných od konstantního trendu	74
17	Dovoz s konstantním trendem, modelem skrytých period a predikcí	75
18	Periodogram dovozu očištěného od polynomického trendu 3. řádu	76
19	Model skrytých period určený Fisherovým testem dat dovozu bez po- lynomického trendu 3. řádu a jeho predikce	78
20	Model skrytých period určený Siegelovým testem dat dovozu bez po- lynomického trendu 3. řádu a jeho predikce	79
21	Dovoz s polynomickým trendem 3. řádu, modelem skrytých period určený Siegelovým testem a predikcí	80
22	Periodogram dat ztrát očištěných od lineárního trendu	82
23	Model skrytých period určený Fisherovým testem dat ztrát očiš- těných od lineárního trendu a jeho predikce	83
24	Model skrytých period určený Siegelovým testem dat ztrát očiště- ných od lineárního trendu a jeho predikce	84
25	Ztráty s lineárním trendem, modelem skrytých period určeným Siegelovým testem a predikcí	85
26	Periodogram dat výroby očištěné od konstantního trendu	86
27	Model skrytých period určený Fisherovým testem u dat výroby očištěných od konstantního trendu a jeho predikce	87
28	Model skrytých period určený Siegelovým testem dat výroby očiš- těných od konstantního trendu a predikce	88

29	Výroba s konstantním trendem, modelem skrytých period určeným Siegelovým testem a predikcí	89
30	Aditivní dekompozice časové řady dovozu pro konstantní trend s modelem skrytých period a náhodnou složkou	93
31	ACF a PACF MODELU 1	94
32	Ověření modelu $MA(1)$ MODELU 1	95
33	Ověření modelu $AR(1)$ MODELU 1	97
34	Predikce z modelu $MA(1)$ MODELU 1	98
35	Aditivní dekompozice časové řady dovozu pro polynomický trend 3. řádu s modelem skrytých period a náhodnou složkou	99
36	ACF a PACF MODELU 2	100
37	Ověření modelu $MA(1)$ MODELU 2	101
38	Predikce z modelu $MA(1)$ MODELU 2	102
39	Aditivní dekompozice dat ztrát pro lineární trend s modelem skrytých period určeným Siegelovým testem a náhodnou složkou . . .	104
40	ACF a PACF MODELU 3	105
41	Aditivní dekompozice dat ztrát pro lineární trend s modelem skrytých period určeným Fisherovým testem a náhodnou složkou . . .	106
42	ACF a PACF MODELU 4	106
43	Ověření modelu $MA(1)$ MODELU 4	107
44	Predikce z modelu $MA(1)$ MODELU 4	108
45	Aditivní dekompozice časové řady výroby pro konstantní trend s modelem skrytých period a náhodnou složkou	109
46	ACF a PACF MODELU 5	110
47	Ověření modelu $ARMA(1, 1)$ MODELU 5	111
48	Predikce z modelu $ARMA(1, 1)$ MODELU 5	112

Seznam tabulek

1	Odhad $\hat{\sigma}(\bar{y})$	47
2	Popis tvaru r_k a r_{kk} v modelech	47
3	Počáteční odhady parametrů a omezení na r_k	48
4	Aproximace σ odhadnutých parametrů	53
5	Určování trendu dovoz	63
6	Určování trendu ztráty	67
7	Určování trendu výroby	69
8	Fisherův test u dat dovozu bez konstantního trendu	73
9	Přesnost odhadů parametrů modelu skrytých period u dovozu očištěného od konstantního trendu	74
10	Predikce dat dovozu s konstantním trendem a modelem skrytých period	76
11	Fisherův test u dat dovozu bez polynomického trendu	77
12	Přesnost parametrů modelu skrytých period podle Fisherova testu u dovozu bez polynom. trendu 3. řádu	77
13	Přesnost parametrů modelu skrytých period určených Siegelovým testem u dat dovozu očištěných od polynom. trendu 3. řádu	79
14	Predikce dat dovozu s polynomickým trendem a modelem skrytých period	80
15	Porovnání určených modelů u dat dovozu	81
16	Fisherův test u dat ztrát bez lineárního trendu	81
17	Přesnost odhadů parametrů modelu skrytých period určených Fisherovým testem u ztrát očištěných od lineárního trendu	82
18	Přesnost odhadů parametrů modelu skrytých period určených Siegelovým testem dat ztrát očištěných od lineárního trendu	83
19	Predikce dat ztrát s lineárním trendem a modelem skrytých period	85
20	Fisherův test u dat výroby bez konstantního trendu	86
21	Přesnost odhadů parametrů modelu skrytých period určených Fisherovým testem výroby očištěné od konstantního trendu	87
22	Přesnost odhadů parametrů modelu skrytých period určených Siegelovým testem výroby očištěné od konstantního trendu	88
23	Predikce dat výroby s konstantním trendem a modelem skrytých period	89
24	Model $ARMA(1, 1)$, $AR(1)$ a $MA(1)$ MODELU 1	95
25	Předpověď z modelu $MA(1)$ MODELU 1	98
26	Model $ARMA(1, 1)$, $AR(1)$, $MA(1)$ MODELU 2	100
27	Předpověď z modelu $MA(1)$ MODELU 2	102
28	Porovnání nalezených modelů náhodných složek dat dovozu	103

29	Model $MA(1)$ MODELU 4	107
30	Predikce z modelu $MA(1)$ MODELU 4	108
31	Model $ARMA(1, 3)$ MODELU 5	110
32	Model $ARMA(1, 1)$ pro data z MODELU 5	111
33	Predikce z modelu $ARMA(1, 1)$ MODELU 5	113
34	Charakteristiky kvality modelů pro data dovozu	114
35	Charakteristiky kvality modelů pro data ztrát	114
36	Charakteristiky kvality modelů pro data výroby	115
37	Porovnání nalezených modelů náhodných složek dat dovozu	115
38	Model $MA(1)$ MODELU 4	116
39	Model $ARMA(1, 1)$ MODELU 5	116

Literatura

- [1] Arlt, J.: Moderní metody modelování ekonomických časových řad, Grada Publishing, Praha, 1. vydání, 1999
- [2] Arlt, J., Arltová, M.: Ekonomické časové řady, Grada Publishing, Praha, 1. vydání, 2007
- [3] Arlt, J., Arltová, M.: Příklady z analýzy ekonomických časových řad, Vysoká škola ekonomická v Praze, Praha, 1. vydání, 1997
- [4] Arlt, J., Arltová, M., Rublíková, E.: Analýza ekonomických časových řad s příklady, Vysoká škola ekonomická v Praze, Praha, 2002
- [5] Cipra, T.: Analýza časových řad s aplikacemi v ekonomii, Nakladatelství technické literatury, Praha, 1. vydání, 1986
- [6] Cipra, T.: Finanční ekonometrie, Nakladatelství Ekopress, Praha, 1. vydání, 2008
- [7] Data dostupná [online], dostupné z: <http://www.eru.cz/cs/elektrina/statistika-a-sledovani-kvality/mesicni-zpravy-o-provozu> [citováno od 20.2.2013 do 14.10.2015]
- [8] Douglas, C.M., Lynwood, A. J., John, S. G.: Forecasting and time series analysis, USA, 2. vydání, 1976
- [9] Energetický regulační úřad [online], dostupné z: <http://www.eru.cz/cs/o-uradu> [citováno 14.1.2015]
- [10] Forbelská, M.: Stochastické modelování jednorozměrných časových řad, Masarykova univerzita, Brno, 1. vydání, 2009
- [11] Hindls, R., Hronová, S., Seger, J., Fischer, J.: Statistika pro ekonomy, Professional Publishing, Praha, 7. vydání, 2006
- [12] Kubáček, L., Kubáčková, L.: Statistika a metrologie, UP v Olomouci, 1. vydání, 2000
- [13] Siegel, A. F.: Testing for periodicity in a time series, 1978 [online], dostupné z: <http://www.dtic.mil/dtic/tr/fulltext/u2/a057350.pdf> [citováno 14.10.2015]
- [14] Wonnacott, T. H., Wonnacott, R. J.: Statistika pro obchod a hospodářství, Victoria Publishing, Praha, 1993