

UNIVERZITA PALACKÉHO V OLMOUCI

PŘÍRODOVĚDECKÁ FAKULTA

KATEDRA MATEMATICKÉ ANALÝZY A APLIKACÍ MATEMATIKY

DIPLOMOVÁ PRÁCE

Podílové fondy z pohledu časových řad



Vedoucí diplomové práce:
Doc. RNDr. Eva Fišerová, Ph.D.
Rok odevzdání: 2012

Vypracoval:
Bc. Jan Večeřa
AME, II. ročník navazující

Prohlášení

Prohlašuji, že jsem tuto diplomovou práci vypracoval samostatně za odborného vedení Doc. RNDr. Evy Fišerové, Ph.D. Dále prohlašuji, že veškeré podklady, které byly použity k vypracování této diplomové práce, jsou uvedeny v seznamu literatury.

V Olomouci dne

Jan Večeřa

Poděkování

Rád bych touto cestou vyjádřil srdečné poděkování Doc. RNDr. Evě Fišerové, Ph.D. za odborné vedení, čas a trpělivost, kterou mi při zpracování této diplomové práce věnovala. Doc. Fišerové patří má úcta a veliký dík za takto vynaloženou ochotu pomoci dovést práci do zdárného konce.

ÚVOD	6
EKONOMICKÁ ČÁST	8
1 Podílové fondy.....	8
1.1 Základní charakteristika podílového fondu.....	9
1.2 Typologie fondů.....	10
1.3 Výhody a nevýhody investic do podílových fondů.....	12
1.4 Podílový list.....	13
MATEMATICKÁ ČÁST	15
2 Časové řady.....	15
2.1 Analýza časové řady.....	16
2.2 Dekompozice časové řady.....	17
2.3 Predikce v časových řadách.....	19
2.4 Hledání trendové funkce.....	21
2.4.1 Metoda nejmenších čtverců.....	23
2.5 Model skrytých period.....	24
2.5.1 Periodogram.....	26
2.5.2 Predikce v modelu skrytých period.....	27
2.6 Box-Jenkinsova metodologie.....	30
2.6.1 Autokovarianční a autokorelační funkce.....	32
2.6.2 Parciální autokorelační funkce.....	34
2.6.3 Lineární proces.....	35
2.6.4 Proces klouzavých součtů MA.....	36
2.6.5 Autoregresní proces AR.....	38
2.6.6 Smíšený proces.....	40
2.6.7 Fáze zpracování modelu.....	41
2.6.8 Předpovědi v Box-Jenkinsově metodologii.....	42

PRAKTICKÁ ČÁST	46
3 Analýza předpovědí	44
3.1 Dekompoziční přístup k PF1.....	48
3.2 Dekompoziční přístup k PF2.....	54
3.3 Box-Jenkinsova metodologie PF1	58
3.3.1 Ověřování stacionarity a identifikace modelu.....	59
3.3.2 Odhad parametrů modelu.....	60
3.3.3 Diagnostika modelu.....	60
3.3.4 Bodová a intervalová predikce.....	63
3.4 Box-Jenkinsova metodologie PF2.....	65
3.4.1 Ověřování stacionarity a identifikace modelu.....	66
3.4.2 Odhad parametrů modelu.....	67
3.4.3 Diagnostika modelu.....	68
3.4.4 Bodová a intervalová predikce.....	70
4 Porovnání se skutečně dosaženými hodnotami	72
4.1 Box-Jenkinsova metodologie.....	72
4.2 Dekompoziční přístup.....	74
4.3 Závěrečné shrnutí.....	76
ZÁVĚR	77
LITERATURA	79
PŘÍLOHA A	80
PŘÍLOHA B	95

ÚVOD

„Vezme-li si investor na pomoc profesionály, vydělá jen 80% toho, co by dokázal sám“

Warren Buffet

V posledních letech se investice do nejrůznějších tržních instrumentů staly poměrně oblíbené. V první řadě už se nejedná pouze o přístup investorů, jak vydělat nadměrné množství finančních prostředků promyšleným alokováním volného kapitálu, ale také o způsob zachování hodnoty finančních prostředků. Investice tedy v dnešní době představují i způsob uložení těchto finančních prostředků se 100 % zajištěním. Pokud budeme uvažovat možnost první, je zcela na místě zaobírat se technikami a metodami, které by riziko, jež podstupujeme, eliminovali ještě o trochu více. Tyto techniky a metody matematických výpočtů dopomáhají investorům, analytikům a pozorovatelům vyhodnotit míry rizika spojená s investicí do konkrétních instrumentů predikovaním jejich budoucích vývoje. Pro svoji diplomovou práci jsem vybral dva konkrétní podílové fondy pro přímou aplikaci matematických přístupů k predikování jejich budoucího vývoje spolu se stanovením predikčních intervalů.

Cílem této práce je popsat matematické přístupy k predikování budoucího vývoje konkrétní ekonomické veličiny (hodnota podílového listu v Kč) a nalézt optimální přístup k modelování vybraných, vysoce volatilních (kolísavých) časových řad v podobě historických hodnot podílových listů, jejichž chování je v čase různé. Vysoká volatilita značí citlivost vývoje na nejrůznější faktory od špatného počasí až po bankrot zcela neznámé společnosti v Asii. Tyto faktory jsou viditelné prakticky denně a hodnoty nejen podílových listů na tyto skutečnosti vždy určitým způsobem reagují.

Teoretická část této diplomové práce bude zaměřena na seznámení se se základními aspekty kolektivního investování a legislativní úpravu této problematiky. Budou vysvětleny důvody pro tento druh investování spolu s jednotlivými typy podílových fondů, které tvoří základní prvek v kolektivním investování. Dále zde budou podrobně popsány

a odvozeny procesy matematického modelování užití ke stanovení budoucího vývoje vybraných podílových fondů.

V praktické části, jež tvoří stěžejní část celé práce, budou tyto fondy zkoumány z pohledu časových řad na základě části teoretické. Při tomto zpracování bude použit software R a MS Excel. Výstupem jednotlivých metod a přístupů bude predikce (jak bodová, tak intervalová) budoucího vývoje těchto fondů. Tyto predikce budou společně se skutečně dosaženými hodnotami zobrazeny v grafech a budou vizuálně porovnány vlastnosti jednotlivých predikcí.

Ekonomická část – převzato z [15]

1 Podílové fondy

Podle Steigaufa je na finančním trhu možno investovat dvojím způsobem, a to přímo či nepřímo, viz [2]. Tato práce je orientována na investování nepřímé, tj. investování s využitím společností, které se specializují na zprostředkování investic na finančním trhu.

Dle [3] podílové fondy úspěšně čelí dvěma zásadním nevýhodám, které jsou:

- *bohatí jednotlivci a instituce měli vždycky přístup k profesionálním správcům peněz*
- *zároveň mohou tyto peníze vhodně rozložit – diverzifikovat (menší sumy neumožňují dostatečně rozložit investice a většina top manažerů v oblasti správy finančních aktiv požaduje poměrně značnou minimální investiční sumu- několik set tisíc až milion korun)*

Většina podílových fondů totiž poskytuje možnost investovat peníze již od několika stovek až tisíc korun a portfolia fondů obsahují desítky až stovky různých investic, zejména cenných papírů.

Myšlenka, která stála u zrodu fondů, je velmi jednoduchá: lidé s podobnými cíli investují své peníze do společného fondu. Profesionální správce-manažer fondu vybere investice v naději, že tyto nejlépe pomohou dosáhnout cíle, které jsou uvedeny ve statutu fondu. Investoři se podílejí proporcionálně na příjmech a ziscích, resp. ztrátách, které tyto fondové investice přinášejí. Protože investoři mohou nakupovat a prodávat podíly ve fondu každý obchodní den, podílové fondy se nazývají také otevřené podílové fondy. Každý podílový fond má portfolio manažera nebo investičního poradce, který řídí investování aktiv fondu podle investičních cílů fondu. Mezi nejznámější kategorie cílů patří, viz [2]:

1. Jistota, zabezpečení hodnoty majetku – snad každý investor požaduje určitou bezpečnost, která zajistí eliminaci rizika.
2. Výnos – jasný důvod, proč do světa investic lidé vstupují, je vytvořit určitý zisk, tzn. Zhodnotit své vložené prostředky.
3. Dlouhodobý růst majetku – ve velké většině fondů se jedná o dlouhodobější investování v horizontu 4 či 5 let.

1.1 Základní charakteristika podílového fondu

Kolektivní investování je založeno na myšlence sdružení finančních prostředků jednotlivých drobných investorů. V [1] Liška uvádí, že je možné ho přirovnat např. ke společné výstavbě bytového domu, kdy stavebníci také sdružují své připravené prostředky a využívají společného postupu převážně za účelem snížení nákladů. U instrumentů kolektivního investování je tomu podobně. Schiller, viz [4], definuje jako hlavní motiv to, aby správa spolčených prostředků dosáhla co největšího efektu, diverzifikace a minimalizace rizik a také proražení či získání přístupu na jiné trhy, obvykle vyhrazené jiným investorům. Princip kolektivního investování je velice starý a má mnoho různých podob. Gladiš ve své publikaci [5] říká: „Mohl bych použít okřídleného úsloví „už staří Římané“, protože už dávno před naším letopočtem se doporučovalo společné podnikání, které omezí případnou ztrátu jedince. Kolektivní investování probíhá na kapitálovém trhu. Podstatou kolektivního investování jakožto podnikatelské činnosti je shromažďování peněžních prostředků od předem neurčitého a neomezeného okruhu právnických a fyzických osob (tedy od veřejnosti) za účelem jejich použití na podnikání na základě principu rozptýlení rizika způsoby, jež však obvykle poskytují širší a komplexnější spektrum služeb. Nicméně základní osnova jejich podnikání i zákonné úpravy je shodná.

1.2 Typologie fondů

Z právního pohledu je možné tyto instituce kolektivního investování rozdělit na tři základní typy (převzato z [1]):

- Otevřené s právní subjektivitou – charakter akciových společností, tzn., že investor se stává akcionářem a svá práva může uplatnit na valné hromadě. Ve světě používána v USA, Velké Británii, Francii.
- Otevřené bez právní subjektivity – spravované investičními společnostmi, které oddělují majetek klientů od svého. Investor není akcionářem, nýbrž podílníkem s podílem na majetku, nicméně bez hlasovacích práv. Tato forma se používá např. v České republice, u některých francouzských a lucemburských fondů (tzv. fondy FCP)
- Uzavřené – je přesně omezen počet emitovaných cenných papírů, investoři nemají právo zpětného odkupu do stanoveného data. I s těmito cennými papíry se na sekundárním trhu obchoduje, obvykle mají nižší likviditu. Toto institucionální ustanovení se používá jen pro některé typy investic, převážně pro dlouhodobé investice typu nemovitostí, případně pro některé typy speciálních fondů-garantovaná portfolia, finanční deriváty atd. Podle legislativy platné v ČR musí být v povolení k vydávání podílových listů uzavřeného podílového fondu stanoven minimální počet vydaných podílových listů a doba, v průběhu které budou podílové listy vydávány.

Členění fondů podle rizika jednotlivých investic, viz [5]:

1) *Fondy peněžního trhu* investují zejména do krátkodobých, vysoce bezpečných dluhopisů tak, aby jejich průměrná doba splatnosti byla kratší než jeden rok. Pro své investování mohou rovněž využít termínovaných vkladů u bank. Kolísání kursu bývá zpravidla nízké, za vysokou míru bezpečí však investor platí relativně nízkým dlouhodobým výnosem. Tyto fondy se hodí pro krátkodobé investice a úložky peněz.

2) *Fondy dluhopisové (obligační)* investují hlavně do dluhopisů, podíl akcií nesmí překročit 10%. Dlouhodobé zhodnocení bývá obvykle vyšší než u fondů peněžního trhu, ale nižší než u akciových fondů. Jsou vhodné pro střednědobý investiční horizont.

3) *Fondy akciové* mají minimálně 66% svého majetku investováno do akcií. Vyznačují se výrazným kolísáním kurzu, třeba i desítek procent v několika měsících. Jsou to fondy velice rizikové a z tohoto důvodu se doporučuje dodržovat investiční horizont 5 let.

4) *Fondy smíšené (balancované)* stojí mezi fondy akciovými a dluhopisovými. Investují do akcií, dluhopisů a nástrojů peněžního trhu současně a manažer portfolia přizpůsobuje v rámci statutu rozložení investic aktuální situaci na trhu. Investiční politika některých balancovaných fondů je velmi flexibilní, čímž však může být pro investora málo předvídatelná. Tyto fondy mají střední až vysoké investiční riziko a tomu odpovídající dlouhodobé zhodnocení. Jsou vhodné pro delší investiční horizont.

5) *Fondy fondů* jsou fondy, které neinvestují přímo do akcií nebo dluhopisů, ale vytvářejí portfolio vytvořené ze specializovaných fondů. Investice do podílových listů jiných fondů znamená pro podílníky lepší rozložení rizika spojeného s investováním.

6) *Garantované fondy* jsou fondy, které garantují investorovi návratnost části nebo celé investice někdy i minimální zaručený výnos v přesně daném investičním období (většinou několik let). Výhodou je omezení možnosti ztráty, nevýhodou delší doba vázání investičních prostředků

Tabulka 1.1 - Shrnutí členění fondů a jejich základní charakteristiky

Typ fondu	Riziko	Minimální investiční horizont	Kurs	Dlouhodobý možný výnos
peněžního trhu	minimální	6 měsíců	stabilní	nízký
dluhopisové	malé až střední	2 roky	mírně kolísavý	nízký až střední
smíšené	střední až vysoké	3 roky	kolísavý	vysoký
akciové	vysoké	5 let	velmi kolísavý	vysoký

Zdroj: autor

1.3 Výhody a nevýhody investic do podílových fondů

Výhody - obecně můžeme říci, že výhody kolektivního investování jsou v tom, že investor nemusí mít žádné znalosti v jednotlivých oborech podnikání, neboť vloží své peněžní prostředky zvláštnímu druhu obchodní společnosti. Kolektivní investování tak snižuje riziko i tím, že podléhá zvláštnímu doзору a musí probíhat v souladu se zákony a na základě licenčního povolení. Tyto skutečnosti potom na vyspělých kapitálových trzích podstatně snižují riziko, že se investor stane obětí neetického jednání. Dalšími výhodami jsou transformační náklady, což jsou náklady na koupi cenných papírů ve formě poplatků makléřské firmě za zprostředkování nákupu. V neposlední řadě patří k výhodám kolektivního investování i odlišný způsob zdaňování či možnost ovlivňovat řízení akciových společností. Jako další očividný argument pro využívání služeb investičních společností je často uváděn fakt, že na rozdíl od individuálního investora má investiční společnost k dispozici moderní systémy, kterými může v podstatě nepřetržitě sledovat akcie, výkyvy a jiné pohyby na světových finančních trzích. Úspěšnost jejich portfolio manažerů ve vyhledávání a realizování úspěšných investic by proto měla být nesrovnatelná oproti jednotlivci, který investuje samostatně.

Nevýhody spočívají v tom, že je nutné platit správní poplatky fondů, což je poplatek za správu kapitálu. Investor rovněž ztrácí možnost ovlivňovat výrazněji zaměření investic - jde o tzv. ztrátu investiční volnosti. Samozřejmě stále zde existuje riziko plynoucí už jenom ze skutečnosti, že jde o investování na kapitálové trhu a i přes limity dané zákonem se mohou vyskytnout společnosti, které nemají s investory čestné úmysly, nebo se stane něco, co nelze předvídat, neboť i kapitálový trh je trhem a ceny se zde určují na základě nabídky a poptávky.

Souhrnně můžeme psát takto, kdy znaménko „+“ značí výhody a znaménko „-“ značí nevýhody:

+ investor nemusí mít žádné znalosti z investičních teorií

správce obhospodařuje podílový fond

+ dochází k diverzifikaci (rozptýlení) rizika

vzhledem k velkému počtu investorů do fondu dochází k rozložení rizika mezi tyto investory

+ nižší náklady na operace s aktivy, výhodnější zdanění

výsledná hodnota fondu, která je investována, je „velká“

+ lepší využití informací a moderních technologií

toto je obecný trend, který je využíván k co největšímu zhodnocení investic

+ větší likvidita investovaných prostředků (otevřené podílové fondy)

likvidita je zajištěna povinností správce odkoupit podíl od investora v případě takového požadavku

+ zajištění prostředků proti ztrátám z neetických praktik

(důsledná regulace subjektu kolektivního investování)

- nutnost platit správní poplatky fondům

- malá možnost ovlivnit zaměření investic

- obecná rizika plynoucí z investování na kapitálovém trhu

1.4 Podílový list

Každý investor obdrží za své peněžní prostředky podílové listy, jejichž prostřednictvím se podílí na všech investicích fondu. Kurz, za který investor podílové listy nakupuje či prodává, odpovídá hodnotě majetku, který připadá na jeden podíl. Z tohoto důvodu nejsou většinou podílové listy obchodovány na veřejných trzích (Burza cenných papírů Praha, RM-System). Není totiž třeba hodnotu podílového listu stanovovat na základě nabídky a poptávky investorů, ale kurz se přesně vypočte z hodnoty majetku. Jednoduše můžeme říci, že podílový list určuje část hodnoty majetku, kterou investor vlastní.

Co se týká práv spojených s podílovým listem, tak je podílový list zvláštním druhem cenného papíru. Majitel podílového listu má právo na odpovídající podíl majetku v podílovém fondu a právo podílet se na výnosu z tohoto majetku podle statutu podílového fondu. Základním právem podílníka je právo na podíl z majetku v podílovém fondu, čili

podílník je spolumajitelem majetku v podílovém fondu. Majetek podílníka roste (nebo klesá) stejným poměrem, jakým se zvyšuje hodnota majetku v podílovém fondu. Podílník má právo na vyplacení svého podílu, který se rovná hodnotě podílového listu, z majetku podílového fondu, jen pokud se podílový fond uzavírá-u podílových listů otevřených podílových fondů i při uplatnění práva na zpětný prodej podílových listů. Podílník má rovněž právo na výnos z majetku v podílovém fondu. Druh a způsob jeho vyplacení je stanoven ve statutu podílového fondu. Výnos může být celý ponechán v majetku podílového fondu, nebo může být celý vyplacen podílníkům, případně může být vyplacen z části a z části ponechán v majetku v podílovém fondu. Vyplacený podíl z majetku v podílovém fondu podléhá zdanění, a to podle zákona č. 268/1992 Sb. Podílník nemá právo na výnos z majetku investiční společnosti, nemá ani právo účastnit se řízení investiční společnosti či ovlivňovat její investiční politiku. Tím se zásadně liší postavení podílníka od postavení akcionáře. Postavení podílníka se tudíž blíží postavení majitele dluhopisu, neboť ani ten nemá právo podílet se na řízení činnosti emitenta dluhopisu. Ovšem majitel dluhopisu má ve vztahu k emitentovi dluhopisu postavení věřitele a má vždy právo na proplacení nominální hodnoty dluhopisu a na výnos z dluhopisu. Podílník nemá vůči investiční společnosti právo na proplacení nominální hodnoty dluhopisu. Pouze při uzavření podílového fondu mu společnost proplácí skutečnou hodnotu jeho podílu. Podílník otevřeného podílového fondu má vůči investiční společnosti právo na zpětný prodej podílového listu, pokud investiční společnost zpětný prodej dočasně nepozastavila.

Podílové listy je možné převádět i na jinou osobu. Způsob převodu závisí na tom, zda byl podílový list vydán na jméno, nebo na doručitele.

Podílový list na jméno je vlastně cenným papírem na řad, tj. jedná se o cenný papír, který se, v případě listinné podoby, převádí rubopisem (přepisem na rubu cenného papíru). Převoditelnost podílového listu rubopisem nemůže investiční společnost ve statutu podílového fondu nijak omezit. Podílový list vydaný na doručitele se převádí předáním z ruky do ruky (tradicí).

Matematická část

2 Časové řady

Časová řada je soubor dat (výsledky experimentů, měření tlaku, počty úspěšných absolventů, počty prodaných výrobků v jednotlivých letech, vývoj hladiny smogu na území ČR apod.), které jsou uspořádány v času ve směru minulost-přítomnost. Všechny výše uvedené časové řady jsou zatížené nejistotou, a proto se nazývají statistické (stochastické). Oproti časovým řadám deterministickým nelze řadu konstruovat na základě funkčního předpisu

$$y_t = f(t)$$

kde y_t značí hodnoty veličin, které modelujeme a t značí časovou proměnnou nabývající hodnot $t = 1, 2, \dots, n$, n značí rozsah časové řady. Tento model nebere v úvahu nejistotu. Proto je nutné stochastický model přepsat takto

$$y_t = f(t) + \varepsilon_t$$

kde ε_t značí náhodnou složku (bílý šum), s nulovou střední hodnotou a nekorelovanými jednotlivými hodnotami. Můžeme tedy říci, že statistické časové řady jsou realizací náhodného procesu.

Podstatnou klasifikací časových řad je dělení na časové řady neekvidistantní, kde jsou jednotlivá data získaná v různých časových intervalech t_i , pro $i = 1, 2, \dots, n$, tj. $t_1 = 2; t_2 = 5; t_3 = 3$ atd. Oproti tomu časová řada ekvidistantní pracuje s daty získanými při stejně dlouhých časových intervalech t .

V této diplomové práci budu pracovat pouze s časovými řadami ekvidistantními (např. roční, čtvrtletní, měsíční pozorování), tj. $t = 1, 2, \dots, n$. Celou časovou řadu potom lze zapsat takto y_t $t=1$ n , tj. y_1, y_2, \dots, y_n . Volba délky intervalu je závislá na pozorované veličině, resp. na jejím charakteru. Např. hodnoty zisků sledované každý pracovní den jsou

relativně nesmyslné a nic neříkající vzhledem k faktu, že zisk je dlouhodobější veličinou. Avšak hodnoty z konce čtvrtletí či pololetí by dlouhodobější vývoj zachycovat mohly.

Časovou řadu tedy lze chápat jako matematickou posloupnost náhodných veličin realizovaného náhodného (stochastického) procesu. Z teoretického hlediska je podstatné rozlišit pojmy „napozorovaná hodnota” a „náhodná veličina”, kdy napozorovaná hodnota je konkrétní realizací náhodné veličiny.

2.1 Analýza časové řady

Účelem analýzy časové řady je především sestavení takového matematického aparátu, který by patřičně vystihoval chování dané řady, tedy jakým způsobem dochází ke generování jednotlivých dat časové řady (členů posloupnosti). Pokud se nám podaří takovýto matematický aparát najít, lze zkonstruovat příslušné budoucí předpovědi (predikce) vývoje dané časové řady. Právě toto je hlavním důvodem, proč matematický aparát vůbec hledat. Takovéto predikce lze využít k optimalizačním procesům, příp. jako podpora v procesu rozhodování. Jako důležitý pomocný nástroj analýzy časových řad je grafické znázornění dané řady. Tato vizuální analýza může již při prvním ohledání řady dát poměrně cenné a důležité informace, které při dalších krocích analýzy mohou zkrátit celý proces jejího zkoumání. Z grafické projekce řady lze na první pohled vyčíst např. dlouhodobý trend časové řady, nebo výrazné periody spolu s frekvencemi, ve kterých se dané významné výkyvy objevují. Dále můžeme také vyčíst, jak stabilní daná časová řada je, případně jakých amplitud dosahují fluktuace. V případě přístupu Boxe-Jenkinse je přímo nezbytné pracovat s grafickými záznamy dané řady z toho důvodu, že daná časová řada musí vykazovat stacionární chování (bude popsáno dále v textu).

V minulosti nebylo možné využívat nástrojů analýzy tak rozšířeně, jako je tomu dnes. Bylo tomu tak převážně proto, že výpočetní technika nedosahovala takové úrovně, s jakou se můžeme setkat v dnešní době. Bez výpočetní techniky by převážná většina analýz trvala řádově několik dnů, týdnů, příp. by nebylo vůbec možné danou časovou řadu analyzovat. Časové řady mohou být velice rozsáhlé a tedy i matice, se kterými se v analýze pracuje, dosahují velkých rozměrů.

Před provedením samotné analýzy je důležité věnovat čas přípravě dat. Pokud je daný datový soubor příliš rozsáhlý, tedy např. obsahuje data, která lze v analýze ignorovat, je vhodné celou řadu upravit selekcí pouze určité části dostupných dat. Například při analýze budoucího vývoje (tak jak tomu bude i v této diplomové práci) podílových fondů je zcela bezvýznamné využívat komplexní soubor historických dat (tvořící až 3500 denních dat), jelikož by se jednotlivé predikce opět stanovovaly na budoucí dny. Takovéto investice jsou dlouhodobou záležitostí, a proto znát předpovědi na několik dnů do budoucna je zcela irelevantní. Eliminace dat by měla být takového rozsahu, aby nedošlo ke ztrátě důležité informace o dané řadě.

Dále je důležité, aby časová řada byla diskretní povahy, což je podmínka, která z principu nemůže být splněna u všech oblastí pozorování a shromažďování dat. Pokud by se jednalo např. o měření teploty, je zřejmé, že charakter této řady by byl spojitý. Je tedy zapotřebí danou časovou řadu (spojitého charakteru) převést na řadu diskretní. To můžeme provést například měřením teploty v určitých časových okamžicích. Pokud by se jednalo o analýzu poptávky po určitém druhu výrobku, daná diskretizace by se provedla akumulací hodnot za sledované období. Těmito metodami diskretizace však může dojít ke ztrátě důležitých vlastností dané řady, např. k nekorelovanosti jednotlivých dat, jak uvádí Cipra ve své publikaci [6].

2.2 Dekompozice časové řady

První přístup, který bude použit při analýze vybraných podílových fondů je přístup dekompoziční. Základem je rozklad (dekompozice) časové řady do těchto složek, viz [7]:

- 1) trend $T r_t$,
- 2) sezónní složka $S z_t$,
- 3) cyklická složka C_t ,
- 4) reziduální E_t ,

Trendová složka – vystihuje tendenci časové řady z dlouhodobějšího hlediska ve sledovaném období. Tato složka identifikuje např. dlouhodobější růst, či naopak

dlouhodobý pokles. V prvním případě bychom si mohli představit časovou řadu, jejíž data jsou přímo závislá od disponibilního důchodu jednotlivců vystupujících na trhu. Tzn., že čím vyšší jsou důchody, tím vyšší tendence mají lidé nakupovat. V opačném případě, tedy v případě nízkých disponibilních důchodů, by lidé nakupovali méně. Jde tedy o existující impulsy ovlivňující přímo trendovou složku. Je důležité poznamenat, že na tyto impulsy (kterých se může v dané časové řadě projevit hned několik) může být nahlíženo různými způsoby, které budou přímo závislé na odborných dovednostech toho, kdo analýzu provádí, příp. na odvětví, ze kterého je daný datový soubor získán.

Sezónní složka – popisuje periodické změny v časové řadě, které se odehrávají během jednoho kalendářního roku a každý rok se opakují, viz [6]. Prakticky se i z běžného života dají vyvodit tendence spotřebitelů. Např. před nástupem léta se dá vysledovat, že prodej opalovacích krémů bude výrazně vyšší, než v kterémkoli jiném období. Tedy půjde o závislost na ročním období.

Cyklická složka – tato složka popisuje kolísavé (flukтуаční) chování časové řady okolo jejího trendu. Dochází tedy k pravidelným poklesům a nárůstům. Obecně lze říci, že vzdálenost od konkrétního vrcholu k následujícímu dnu, ať už ve směru času, či ve směru sledované veličiny není konstantní.

Reziduální složka – označována také jako „bílý šum (White noise)“ vzniká odstraněním všech zbývajících složek, tedy trendové, sezónní a cyklické (souhrnně nazývané jako periodické). Tato složka je zcela náhodná a charakteristické pro tuto složku je, že střední hodnota je rovna nule a všechny prvky jsou navzájem nekorelované. Tato složka se díky svým vlastnostem nazývá nesystematická, oproti tomu zbylé složky (trendová, sezónní a cyklická) tvoří systematickou část dekompozice.

Matematický popis dekompozice může být dvojího druhu:

- 1) *Aditivní přístup* – používáme konkrétní absolutní hodnoty jednotlivých složek rozkladu, které jsou navíc měřeny ve stejných jednotkách, jako pozorovaná řada y_t ,

$$y_t = T r_t + S z_t + C_t + \varepsilon_t$$

- 2) **Multiplikativní přístup** – u tohoto přístupu se pracuje s absolutními hodnotami složky trendu, ostatní složky jsou uvažovány v relativních hodnotách (jsou tedy bezrozměrné).

$$y_t = Tr_t S z_t C_t \varepsilon_t$$

Nutno poznamenat, že výskyt všech těchto složek v časové řadě najednou nemusí nastat a v praxi často nenastává.

2.3 Predikce v časových řadách

Primární účelem analýzy časové řady je pochopení systému generování jednotlivých dat. Najdeme-li takovýto matematický aparát, jsme schopni zkonstruovat i budoucí předpovědi dané časové řady. Je zřejmé, že v praxi mohou mít tyto předpovědi vysokou hodnotu a mohou ekonomickým subjektům přinést nemalé ušetřené částky, příp. významně navýšit zisky, či naopak zredukovat skladové zásoby a tím i celkové náklady na skladování. Prakticky se konstruují dva druhy předpovědí:

- 1) Bodové předpovědi – tato předpověď představuje náš odhad budoucí hodnoty uvažované časové řady. Bodová předpověď je vždy zatížena chybou (způsobenou odhadem parametrů, redukcí dat apod.), z tohoto důvodu je zapotřebí uvažovat danou předpověď s rezervou. Abychom mohli tuto chybu vyčíslit, je nutné napočítat předpovědní intervaly
- 2) Předpovědní intervaly – jsou de facto analogií intervalů spolehlivosti z matematické statistiky. Tyto intervaly pokrývají danou hodnotu budoucího vývoje s určitou pravděpodobností. Tzn., že předpovědní interval udává dolní a horní mez, mezi kterými se bude daná budoucí hodnota nacházet s pravděpodobností $1 - \alpha$, kde α je zvolená hladina významnosti.

V praxi lze využít dva základní přístupy k predikci časové řady. Oba tyto přístupy se liší přístupností k historickým datům a tedy relativním rozdílem ve složitosti výpočtů. Často využívaným přístupem je přístup kvalitativní (viz [6]), spíše využívaný v managementu rizika ke stanovení nejpravděpodobnější hodnoty vývoje sledovaného kritéria (zisku, celkových nákladů apod.) v budoucnu. V případě nedostupnosti historických dat je vhodné volit tento přístup, v případě opačném, tedy pokud známe historický vývoj sledované řady, použijeme přístupy kvantitativní. Právě těmto přístupům bude věnována patřičná pozornost v dalším textu.

V textu výše již byla zmíněna praktická důležitost intervalových předpovědí. Tato důležitost vyplývá z podstaty stochastického modelu, který je vždy zatížen nejistotou. Právě z tohoto důvodu lze zavést další pojem *chyba předpovědi*. Tato veličina vystihuje absolutní rozdíl mezi skutečným vývojem dané časové řady a jejími predikcemi. Definujeme je tedy takto:

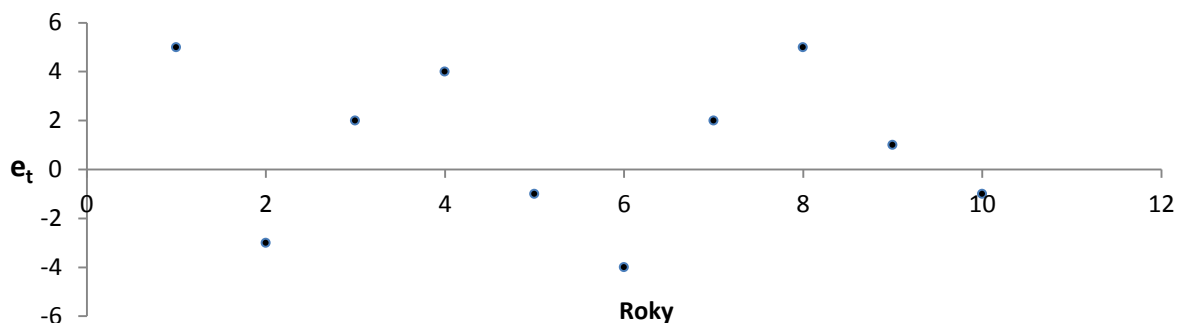
$$e_T = y_T - \hat{y}_T^p$$

kde

e_T	chyba předpovědi v čase T
y_T	skutečná hodnota v čase T
\hat{y}_T^p	predikovaná hodnota v čase T

Je zřejmé, že na velikosti chyby předpovědi se může podílet velké množství faktorů, zejména pak to, jak dobře jsme namodelovali jednotlivé složky v dekompozici. Cipra ve své publikaci [6] píše, že vhodným způsobem eliminace špatných předpovědí je provádět předpovědi již známých dat a tak otestovat, zdali je zvolená technika vhodná či ne. V ideálním případě by rozložení chyb předpovědí mělo průběh na obr. 2.1. Z obr. 2.1 je vidět, že daná řada chyb vykazuje stacionární chování, tj. ustálené pravděpodobnostní chování v čase. Toto chování je charakteristické tím, že střední hodnota $E(e_t) = 0$ a rozptyl $D(e_t)$ je konstantní.

Obrázek 2.1 – Rozdělení chyb předpovědí



Zdroj: Vlastní zpracování

2.4 Hledání trendové funkce

Máme-li k dispozici historická data, je možné matematicky modelovat trend dané řady neboli zjišťovat tendenci vývoje. Obecně se pracuje s předpokladem, že má analyzovaná řada tento tvar

$$y_t = Tr_t + \varepsilon_t.$$

Tzn., že tato časová řada obsahuje pouze trendovou a reziduální složku, případně na tento tvar byla jistými transformacemi převedena. O těchto metodách se dozvíme dále. Na základě tohoto předpokladu můžeme užít regresní analýzy. Většina časových řad z ekonomické praxe může být popsána některou matematickou křivkou, objevují se však případy, kdy s jednoduchou regresní analýzou nevystačíme, a je nutné použít komplikovanější přístup. Tato situace, jak dále uvidíme, nastala i při analýze zvolených podílových fondů. Zmíněná regresní analýza je založena na metodě nejmenších čtverců. Jako měřítko kvality jednotlivých funkcí se používá *index determinace*, *SSE* (Sum of Squared Errors – součet čtverců reziduí), *MSE* (Mean Squared Error – střední čtvercová chyba) aj.

Jednotlivé charakteristiky se počítají dle následujících vzorců

- Index determinace

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

Index determinace vystihuje míru vhodnosti daného modelu vzhledem ke zkoumané časové řadě. Nabývá hodnot z intervalu $\langle 0,1 \rangle$, kde hodnota 0 znamená, že daný model je zcela nevhodný s ohledem na danou časovou řadu a naopak hodnota 1 znamená, že tento model danou řadu výborně vystihuje. Žádoucí jsou tedy hodnoty blížíící se k 1.

- SSE

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

Součet čtverců chyb neboli rozdílů mezi napozorovanými hodnotami a hodnotami odhadnutými. Čím menší hodnoty SSE daný model dosahuje, tím vhodnější tento model je.

- MSE

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

Střední čtvercová chyba [6] udává, jak se v průměru liší čtverce rozdílů napozorovaných a vyrovnaných hodnot. Nižší hodnoty tohoto ukazatele značí kvalitnější model.

- MAPE

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

Střední, absolutní, procentuální chyba měří přesnost procentuálně. Čím nižší hodnoty MAPE, tím kvalitnější jsou naše odhady.

Vzhledem k predikcím je vhodnější užít upravený index determinace (*Adjusted Coefficient of Determination*), viz [12]. Tento přístup odstraňuje vliv počtu regresních parametrů. Na rozdíl od standardního indexu determinace nemusí platit, že tento index roste s vyšším

počtem užitých parametrů v modelu. Upravený index determinace počítáme následujícím způsobem

$$I_{adj}^2 = 1 - \frac{(1 - I^2)(n - 1)}{(n - p)},$$

kde p je počet parametrů v modelu a n značí počet pozorování. Je zřejmé, že s rostoucím počtem pozorování se upravený index determinace blíží standardnímu indexu determinace.

2.4.1 Metoda nejmenších čtverců

Tato metoda je jedna z nejvíce používaných metod pro zpracování relativně nepřesných dat a jejich prokládání (aproximaci). Nyní si odvodíme jeden z důležitých vzorců ve statistice. K odvození tohoto vzorce lze užít mnoho přístupů, zde bude použit přístup založený na ortogonalitě.

Předpokládejme např., že daná časová řada, resp. její trend je dobře popsán lineární funkcí $y_t = b + at + \varepsilon_t$, potom tento vztah můžeme zapsat maticově následujícím způsobem jako $y = X\beta$. Vektor y je sloupcový vektor pozorování o rozměru $n \times 1$, kde n značí počet pozorování. Matici $X_{n \times p}$ nazýváme designovou maticí o rozměru $n \times p$, kde n značí rozsah souboru a p značí počet neznámých parametrů v modelu. Pro náš případ lineárního trendu by designová matice měla rozměr $n \times 2$, jelikož lineární funkce sestává ze dvou parametrů, a měla by tedy následující tvar

$$X = \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ \cdot & \cdot \\ \cdot & \cdot \\ 1 & n \end{pmatrix},$$

kde sloupec jedniček značí výskyt absolutního členu v trendové funkci. Dále označme $\beta_{p \times 1}$ vektor neznámých parametrů, který by opět pro náš příklad lineárního trendu měl

následující tvar $\beta = \begin{pmatrix} b \\ a \end{pmatrix}$. Potom hledáme takový reziduální vektor $e_{n \times 1}$, tedy vektor rozdílů odhadnutých a vypočítaných hodnot, který bude mít minimální euklidovskou délku. Znamená to tedy, že hledaný vektor $e_{n \times 1}$ má být ortogonální na $\mathbb{R}(X)$, tj. obor hodnot.

Platí tedy:

$$X^T e = 0.$$

Dosadíme-li nyní za reziduum, dostaneme

$$X^T (X \hat{\beta} - Y) = X^T X \hat{\beta} - X^T Y = 0$$

a tedy

$$\hat{\beta} = (X^T X)^{-1} X^T Y \quad (2.1)$$

Předpokládáme zde lineární nezávislost sloupců matice $X_{n \times p}$, z čehož přímo plyne jednoznačnost řešení. Pomocí vzorce (2.1) tedy můžeme odhadnout parametry regresního modelu, které dále budeme značit $\hat{\beta}$, kde stříška nad písmenem beta značí odhadnutou hodnotu. Rozptyl tohoto odhadu je jednoduché odvodit ve tvaru

$$\text{var}(\hat{\beta}) = \sigma_\varepsilon^2 (X^T X)^{-1},$$

kde σ_ε^2 odhadujeme dle následujícího vzorce

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

2.5 Model skrytých period

Podaří-li se nám identifikovat vhodnou trendovou funkci (trendovou složku), je nutné zkoumat i složku periodickou (sezónní+cyklickou).

Značným přínosem v analýze podílových fondů z pohledu časových řad by bylo nalezení významných period při zkoumání periodické složky tzv. *spektrální analýzou*. Pokud takovéto významné periody v časové řadě nalezneme, můžeme jednoduše sestavit *model skrytých period*, na základě něhož lze předvídat pravidelné výkyvy, neboli „nálady“ tržních spekulantů a tohoto využít k peněžnímu přilepšení. Při modelování periodických složek je nutné pracovat s časovou řadou vykazující konstantní trend, pokud daná časová řada tento předpoklad nesplňuje, je nutné od této řady odečíst trendovou složku.

Obecně při spektrální analýze považujeme časovou řadu za kombinaci sinusovek a kosinusovek, které mají odlišné frekvence a amplitudy. Základem je tedy tzv. trigonometrický polynom

$$y_t = a_0 + \sum_{j=1}^H a_j \sin \omega_j t + \sum_{j=1}^H b_j \cos \omega_j t + \varepsilon_t, \quad t = 1, 2, \dots, n, \quad (2.2)$$

kde pro n sudé platí, že $H = \frac{n}{2}$ a pro n liché je $H = \frac{n-1}{2}$. Dále ω_j značí j -tou frekvenci, tedy udává počet opakování určitého periodického děje za jednotku času a počítá se jako $\omega_j = \frac{2\pi j}{n}$ pro $j = 1, 2, \dots, H$. Polynomiální koeficienty a_j, b_j , mají funkci amplitud, a čím více se tyto koeficienty liší od nuly, tím jsou dané harmonické kmity významnější. ε_t je reziduální složka. Převrácenou hodnotou frekvence ω_j dostaneme tzv. *Fourierovy periody*, neboli vlnové délky, udávající za jakou dobu proběhne jeden cyklus. Vzhledem k tomu, že ne všechny tyto Fourierovy periody v modelu mohou být považovány za významné, je nutné užít *periodogram* (viz níže) a určit tak statistickou významnost jednotlivých amplitud kmitů, tedy určit takové Fourierovy periody, které budou výrazně odlišné od ostatních. Předpokládejme, že počet takto významných period je r , potom při splnění předpokladu konstantního trendu lze výraz (2.2) zapsat v tomto tvaru

$$y_t = a_0 + \sum_{j=1}^r a_j \sin(\omega_j t) + \sum_{j=1}^r b_j \cos(\omega_j t) + \varepsilon_t, \quad t = 1, 2, \dots, n. \quad (2.3)$$

Do modelu (2.3) jsou již zanašeny pouze ty Fourierovy periody, které byly shledány významnými a tento model nazýváme *model skrytých period*. Tento regresní model je

lineární v parametrech a je proto možné užít metodu nejmenších čtverců pro odhad parametrů a_0, a_j, b_j . Odhady těchto parametrů se tedy počítají následovně, viz [6]

$$\hat{a}_0 = \frac{1}{n} \sum_{t=1}^n \hat{y}_t^{(0)}, \quad \hat{a}_j = \frac{2}{n} \sum_{t=1}^n y_t \sin(\omega_j t), \quad \hat{b}_j = \frac{2}{n} \sum_{t=1}^n y_t \cos(\omega_j t), \quad j = 1, 2, \dots, r \quad (2.4)$$

2.5.1 Periodogram

K určení periodicity, neboli nalezení významných Fourierových period se v běžné praxi užívá periodogramu, který definujeme jako funkci úhlové frekvence nabývající reálných hodnot, viz [6]. Periodogram určuje podíly (příspěvky) jednotlivých frekvencí na celkové variabilitě odhadnuté časové řady. Hodnoty periodogramu počítané podle následujícího vzorce

$$v_j = \frac{1}{2} (\hat{a}_j^2 + \hat{b}_j^2) \quad (2.5)$$

vynášíme pro snadnější orientaci do grafu (též periodogramu) na svislou osu, zatímco na ose vodorovné jsou zachyceny příslušné frekvence. Významnost jednotlivých Fourierových period testujeme Fisherovým testem tak, že hodnoty periodogramu seřadíme od největší po nejmenší a postupně dosazujeme do následujícího vztahu

$$Q_j = \frac{v_j}{\sum_{j=1}^H v_j} \quad (2.6)$$

Ze vztahu 2.6 je zřejmé, že pokud by hodnoty v_j pro $j = 1, 2, \dots, H$ nabývaly přibližně stejných hodnot, potom by se hodnota Q_j blížila $\frac{1}{H}$. V případě, kdy by existovala hodnota v_j výrazně vyšší, potom by se hodnota Q_j blížila jedné. Testová statistika má tedy tvar

$$F = \max_j Q_j \quad \text{pro } j = 1, 2, \dots, H .$$

Testujeme nulovou hypotézu, že daná časová řada neobsahuje významnou periodu, tedy

$$H_0 : y_t = a_0 + \varepsilon_t.$$

Hodnoty testovací statistiky porovnáváme s kritickou hodnotou Fisherova rozdělení $g_F(\alpha)$ na hladině významnosti α . Nulovou hypotézu zamítáme, pokud testovací statistika překročí tuto kritickou hodnotu, tedy pokud $F > g_F$.

Samotný test provádíme v několika krocích:

- 1) Seřadíme hodnoty periodogramu od největší po nejmenší, přičemž hodnotu největší označíme ν_1 a hodnotu nejmenší označíme ν_H .
- 2) Spočítáme hodnotu testovací statistiky a porovnáme s kritickou hodnotou.
- 3) Pokud zjistíme významnost dané periody, je nutné testovat významnost druhé nejvyšší hodnoty spočítané pro periodogram tak, že jmenovatel ve výrazu (2.6) snížíme o hodnotu nejvyšší v periodogramu a tedy hodnotu H snížíme o jednotku. Takto postupujeme dále, dokud nacházíme významné periody.

2.5.2 Predikce v modelu skrytých period

Uvažujme model skrytých period s r významnými periodami ve tvaru

$$y_t = a_0 + \sum_{j=1}^r a_j \sin(\omega_j t) + \sum_{j=1}^r b_j \cos(\omega_j t) + \varepsilon_t, \quad t = 1, 2, \dots, n, \quad (2.7)$$

kde a_0, a_j, b_j pro $j = 1, 2, \dots, r$ jsou parametry trigonometrického polynomu. Odhady těchto parametrů provádíme dle (2.4). Mějme tedy vektor odhadu parametrů

$\hat{\gamma} = (a_0, a_1, \dots, a_r, b_1, \dots, b_r)$. Variance tohoto vektoru je

$$\text{var } \hat{\gamma} = (X^T X)^{-1} \sigma_\varepsilon^2,$$

kde $X_{n \times p}$ je designová matice, která má tvar

$$X_{n \times p} = \begin{pmatrix} 1 & \sin(\omega_{j_1} 1) & \cos(\omega_{j_1} 1) & \sin(\omega_{j_2} 1) & \cos(\omega_{j_2} 1) & \dots & \sin(\omega_{j_r} 1) & \cos(\omega_{j_r} 1) \\ 1 & \sin(\omega_{j_1} 2) & \cos(\omega_{j_1} 2) & \sin(\omega_{j_2} 2) & \cos(\omega_{j_2} 2) & \dots & \sin(\omega_{j_r} 2) & \cos(\omega_{j_r} 2) \\ \cdot & \cdot & \cdot & \cdot & \cdot & \dots & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \dots & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \dots & \cdot & \cdot \\ 1 & \sin(\omega_{j_1} n) & \cos(\omega_{j_1} n) & \sin(\omega_{j_2} n) & \cos(\omega_{j_2} n) & \dots & \sin(\omega_{j_r} n) & \cos(\omega_{j_r} n) \end{pmatrix}.$$

Jelikož je σ_ε^2 neznámé, je třeba ho odhadnout

$$\sigma_\varepsilon^2 = \text{var } \varepsilon_t = \frac{1}{n-1} \sum_{t=1}^n (y_t - \hat{y}_t)^2. \quad (2.8)$$

Abychom mohli dále odvodit predikční intervaly, je zapotřebí odvodit $\text{var}(y_T - \hat{y}_T)$.

Jelikož jsou y_T a \hat{y}_T nezávislé, je jejich kovariance nulová a je tedy zřejmé, že

$$\text{var}(y_T - \hat{y}_T) = \sigma_\varepsilon^2 + 0 + \text{var } \hat{y}_T. \quad (2.9)$$

Dále je nutné odvodit $\text{var } \hat{y}_T$ takto

$$\text{var } \hat{y}_T = \text{var}(x_T^T \hat{\gamma}) = x_T^T \text{var } \hat{\gamma} x_T = \sigma_\varepsilon^2 x_T^T (X^T X)^{-1} x_T, \quad (2.10)$$

kde $X_{n \times p}$ je designová matice a x_T je řádkový vektor matice X pro $T > n$. Dosadíme-li zpětně do (2.9), dostaneme následující výraz

$$\text{var}(y_T - \hat{y}_T) = \sigma_\varepsilon^2 + \sigma_\varepsilon^2 x_T^T (X^T X)^{-1} x_T = \sigma_\varepsilon^2 [1 + x_T^T (X^T X)^{-1} x_T], \quad (2.11)$$

kde σ_ε^2 je neznámé, proto je nutné jej odhadnout dle 2.8. Tedy např. 95%ní predikční interval pro budoucí pozorování v čase T bude vypadat takto

$$\left(\hat{y}_T - \sqrt{\hat{\sigma}_\varepsilon^2 [1 + x_T^T (X^T X)^{-1} x_T]} \cdot t_{n-p}(0,05) ; \hat{y}_T + \sqrt{\hat{\sigma}_\varepsilon^2 [1 + x_T^T (X^T X)^{-1} x_T]} \cdot t_{n-p}(0,05) \right),$$

kde $t_{n-p}(0,05)$ značí kritickou hodnotu studentova rozdělení s $n-p$ stupni volnosti na hladině významnosti $\alpha = 0,05$. Je nutné dodat, že tento odhad lze užít pouze pro periodickou složku. Pokud bychom tímto způsobem počítali predikční intervaly pro celkový model, tedy se zahrnutím i trendové složky, nebyla by nejistota této složky v odhadu zahrnuta. Predikční intervaly by tedy nebyly správně zkonstruované. Proto je na místě, známe-li všechny významné periody a trendovou funkci, zkonstruovat nový, celkový model, který by obsahoval jak trendovou, tak zmíněnou periodickou složku. Počet parametrů, které je zapotřebí odhadnout je p (Trendová+Periodická složka). Designová matice tedy vypadá následovně

$$X = \begin{pmatrix} 1 & t_1 & t_1^2 & \cdot & t_1^m & \sin(\omega_1 t_1) & \cos(\omega_1 t_1) & \cdot & \sin(\omega_r t_1) & \cos(\omega_r t_1) \\ 1 & t_2 & t_2^2 & \cdot & t_2^m & \sin(\omega_1 t_2) & \cos(\omega_1 t_2) & \cdot & \sin(\omega_r t_2) & \cos(\omega_r t_2) \\ 1 & t_3 & t_3^2 & \cdot & t_3^m & \sin(\omega_1 t_3) & \cos(\omega_1 t_3) & \cdot & \sin(\omega_r t_3) & \cos(\omega_r t_3) \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 1 & t_n & t_n^2 & \cdot & t_n^m & \sin(\omega_1 t_n) & \cos(\omega_1 t_n) & \cdot & \sin(\omega_r t_n) & \cos(\omega_r t_n) \end{pmatrix}$$

kde n značí rozsah souboru a r počet nalezených významných period. S touto maticí plánu tedy odhadneme nové parametry celkového modelu a následně spočítáme predikční intervaly dosazením do vztahu pro výpočet predikčních intervalů. Je tedy zřejmé, že i vektor x bude vypadat takto

$$x = \begin{pmatrix} 1 \\ T \\ T^2 \\ \cdot \\ T^m \\ \sin(\omega_1 T) \\ \cos(\omega_1 T) \\ \cdot \\ \sin(\omega_r T) \\ \cos(\omega_r T) \end{pmatrix}^r,$$

kde $T=n+1, n+2, \dots$ jsou predikovaná období, m značí stupeň polynomu trendové funkce, r počet nalezených významných period.

2.6 Box-Jenkinsova metodologie

Box-Jenkinsova metodologie dovoluje modelování pouze stacionárních časových řad, případně takových, které lze na stacionární převést určitou transformací (nejčastěji diferencováním). Další podmínkou pro užití Box-Jenkinsovy metodologie je minimální rozsah zkoumaného souboru, tedy $n > 50$, převzato z [8].

V odborné literatuře se obecně rozlišují dva typy stacionarity (viz [7]):

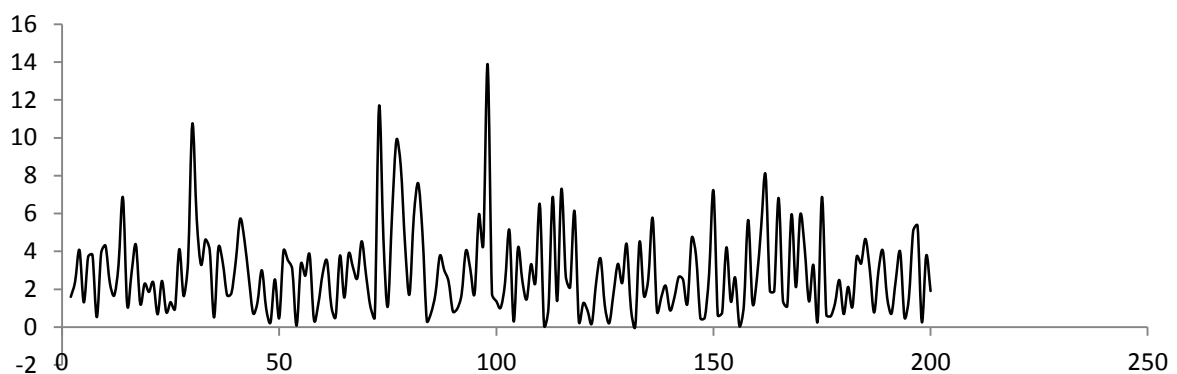
- 1) Striktní stacionarita- je založena na předpokladu, že časová řada, resp. její chování dané jejím rozdělením, je v čase invariantní, tj. nemění se při posunu v čase.
- 2) Slabá stacionarita- vyžaduje ustálené chování náhodného procesu, tj. konstantní střední hodnotu a konstantní rozptyl. Dále slabá stacionarita vyžaduje, aby platilo

$$\text{cov}(y_t, y_s) = \text{cov}(y_{t+h}, y_{s+h})$$

tj. kovariance mezi jednotlivými pozorováními není závislá na umístění těchto pozorování v řadě, ale na jejich vzájemné vzdálenosti.

Striktní stacionarita automaticky implikuje slabou stacionaritu. V této práci budeme uvažovat pouze slabou stacionaritu, jakožto postačující vlastnost pro užití Box-Jenkinsovy metodologie. Pro bližší pochopení slabé stacionarity pomohou následující obr. 2.2 a 2.3:

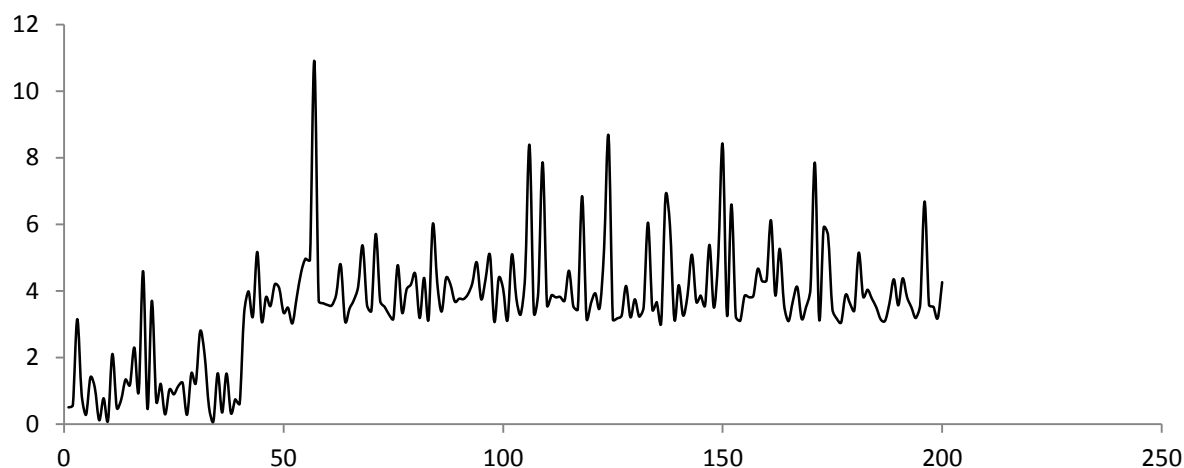
Obrázek 2.2 – Nestacionární časová řada



Zdroj: vlastní zpracování pomocí generátoru náhodných čísel v MATLABu.

Z obr. 2.2 je vidět, že daná časová řada nesplňuje podmínku konstantního rozptylu. Diferencováním této řady bychom stacionarity dosáhli.

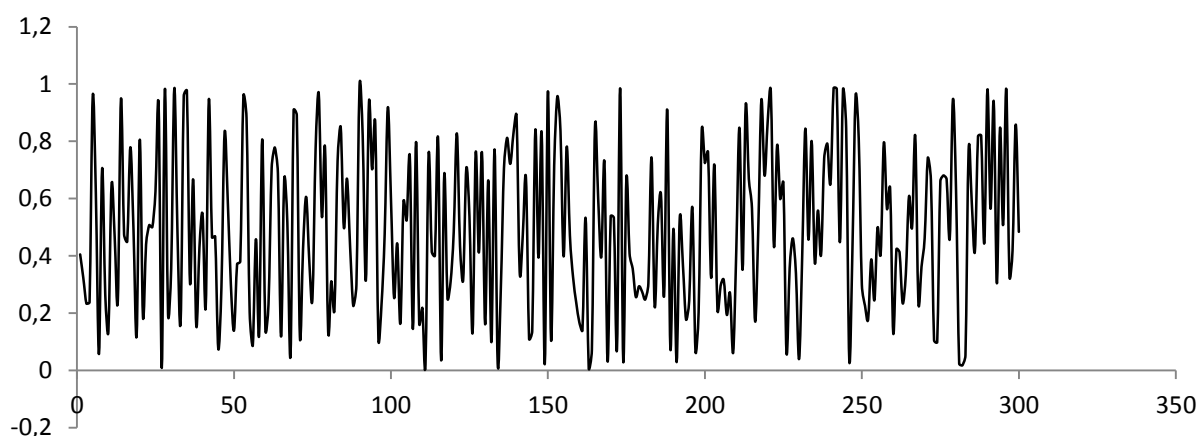
Obrázek 2.3 – Nestacionární časová řada



Zdroj: vlastní zpracování pomocí generátoru náhodných čísel v MATLABu.

Z obr. 2.3 můžeme vypočítat, že tato časová řada se pohybuje na dvou úrovních středních hodnot, a tudíž je porušen předpoklad konstantní střední hodnoty po celé délce časové řady. Opět bychom tuto časovou řadu mohli transformovat pomocí diferencování na časovou řadu stacionární.

Obrázek 2.4 – Stacionární časová řada



Zdroj: Vlastní zpracování pomocí generátoru náhodných čísel v MATLABu.

Na obr. 2.4 jde vidět, že tato řada se již jako stacionární jeví, jelikož po celé své délce $n = 300$ lze vypočítat konstantní střední hodnotu i konstantní rozptyl. Na tomto obrázku je také ilustrován fakt, že vlastnost stacionarity můžeme vypočítat pouhým okem při grafickém znázornění dané časové řady. Jsou však situace, kdy takováto vizuální kontrola nemusí dávat relevantní výsledek a je nutné použít jiné, již statistické metody ke zkoumání stacionarity.

2.6.1 Autokovarianční a autokorelační funkce

Je-li splněna podmínka stacionarity, lze definovat autokovarianční funkci v bodě k jako

$$\gamma_k = \text{cov}(y_t, y_{t+k})$$

a dále také lze definovat autokorelační funkci (označovanou jako ACF)

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \frac{\text{cov}(y_t, y_{t+k})}{\sigma_y^2}, \quad \text{pro } k = \dots, -1, 0, 1, \dots \quad (2.12)$$

kde $\sigma_y^2 = \text{var}(y_t)$ značí rozptyl dané řady. Předpona auto- je vysvětlena dosazováním hodnot z téže časové řady. Vzhledem k tomu, že autokorelační funkce je funkce sudá, můžeme se omezit pouze na $k \geq 0$, přičemž platí $\rho_0 = 1$ a $|\rho_k| \leq 1$ pro $k > 0$, viz [6]. Graf zobrazující hodnoty autokorelační funkce nazýváme *korelogram*.

Průběh autokorelační funkce lze také považovat za indikátor stacionarity či nestacionarity. V případě stacionární časové řady je průběh absolutní hodnoty ρ_k relativně rychle klesající. V případě nestacionární časové řady je pokles velice pomalý.

Dále se z průběhu autokorelační funkce vyvozuje důležitý poznatek spočívající ve stanovení tzv. identifikačního bodu $k = k_0$, tedy takového bodu pro který platí, že

$$\rho_k = 0 \quad \text{pro } \forall k > k_0.$$

Případně můžeme vyslovit hypotézu o neexistenci takového bodu. Právě na základě vyřešení existence a případné hodnoty identifikačního bodu lze provést samotnou identifikaci modelu. Jelikož neznáme skutečnou autokorelační funkci uvažované časové řady, je zapotřebí využít jejich odhadů.

Odhad autokovarianční funkce se obecně provádí dle standardního vzorce pro výpočet kovariance s tím rozdílem, že pracujeme s jedním datovým souborem, který dělíme na dvě části. První část je tvořena původními daty y_t až do okamžiku $n-k$, zatímco druhá část je tvořena postupně daty o k kroků dopředu y_{t+k} pro $k = 0, 1, 2, \dots, n-1$.

Odhad autokovarianční funkce budeme značit c_k . Vzorec pro odhad autokovarianční funkce tedy vypadá následovně

$$c_k = \sum_{t=1}^{n-k} \frac{(y_t - \bar{y})(y_{t+k} - \bar{y})}{n}. \quad (2.13)$$

Odhad autokorelační funkce budeme značit r_k a je tedy dán výrazem

$$r_k = \frac{c_k}{c_0}, \quad k = 0, 1, 2, \dots, n-1. \quad (2.14)$$

Dle Boxe a Jenkinse (viz [12]) je vhodné počítat r_k až po $k = \frac{n}{4}$. Odhady autokorelační funkce obecně slouží ke dvěma účelům:

- 1) zkoumání vlastnosti stacionarity analyzované časové řady
- 2) stanovení identifikačního bodu

Je tedy nutné, abychom byli schopni statisticky ověřit platnost $\rho_k = 0$. K tomuto testování se standardně používá *Bartlettova aproximace*, která de facto při rozhodování o nulové hypotéze $H_0: \rho_k = 0$ porovnává hodnotu r_k vypočítanou pomocí vzorce (2.14), s dvojnásobkem směrodatné odchylky pro tento odhad, tj. s hodnotou $2\sigma(r_k)$. Konkrétní volba hodnoty 2 odpovídá zhruba hladině významnosti testu rovné 0,05. Výraz $\sigma(r_k)$ se aproximuje výrazem, převzato z [7].

$$\sqrt{\frac{1}{n} \left(1 + 2 \sum_{j=1}^{k_0} r_j^2 \right)} \quad \text{pro } k > k_0. \quad (2.15)$$

2.6.2 Parciální autokorelační funkce

Parciální autokorelační funkce (dále jen PACF) se v Box-Jenkinsově metodologii používá stejně jako autokorelační funkce k identifikaci modelu a k jeho následnému ověřování.

V literatuře [6] a [8] se PACF značí symbolem ρ_{kk} , a tato hodnota se definuje jako parciální korelační koeficient y_t, y_{t+k} při pevných hodnotách $y_{t+1}, y_{t+2}, \dots, y_{t+k-1}$, tj. korelační koeficient pro y_t, y_{t+k} očištěný od vlivu veličin $y_{t+1}, y_{t+2}, \dots, y_{t+k-1}$. Platí

$$\rho_{kk} = \frac{|P_k^*|}{|P_k|}, \quad (2.16)$$

kde $|P_k|$ značí determinant matice autokorelací získaných podle 2.14. Matice autokorelací je symetrická, kde prvek na pozici (i, j) značí odhad autokorelace o $k = j - i$ kroků dopředu, pro $i \leq j$. Jelikož $\rho_k = 1$ pro $k = 0$, jsou na hlavní diagonále jedničky. Obecně tedy matice P_k vypadá následovně

$$P_k = \begin{pmatrix} 1 & \rho_1 & \rho_2 & \dots & \dots & \rho_{k-1} \\ \rho_1 & 1 & \rho_1 & & & \rho_{k-2} \\ \rho_2 & \rho_1 & \cdot & & & \\ & & & \cdot & & \\ & & & & \cdot & \\ \rho_{k-1} & \rho_{k-2} & & & & 1 \end{pmatrix}.$$

Předefinováním posledního sloupce této matice tak, že bude tvořen vektorem $\rho_1, \rho_2, \dots, \rho_k^T$, získáme matici P_k^* , viz [8] str. 64. Vzhledem k faktu, že skutečné hodnoty PACF neznáme, je nutné je opět odhadnout. Odhady PACF označme r_{kk} . Výpočet odhadnutých hodnot PACF je dán následujícím vztahem (převzato z [6])

$$r_{kk} = \frac{r_k - \sum_{j=1}^{k-1} r_{k-1,j} r_{k-j}}{1 - \sum_{j=1}^{k-1} r_{k-1,j} r_j}. \quad (2.17)$$

Cipra v [6] uvádí, že tyto vzorce jsou hojně využívány matematickými softwary a obecně počítači při rekurentních výpočtech odhadů. Tyto odhady opět využijeme k tomu, abychom statisticky ověřili domněnku, že $\rho_{kk} = 0$. K tomu se používá *Quenouilleova aproximace* pro směrodatnou odchylku odhadu r_{kk} . Platí, že (převzato z [8])

$$\sigma(r_{kk}) \sim \sqrt{\frac{1}{n}}. \quad (2.18)$$

Pokud absolutní hodnota odhadnuté hodnoty r_{kk} překročí dvojnásobek směrodatné odchylky, nulovou hypotézu $H_0: \rho_{kk} = 0$ zamítáme na hladině $\alpha = 0,05$.

2.6.3 Lineární proces

V oblasti analýzy časových řad je lineární proces definován jako nekonečná řada, která je tvořena hodnotami bílého šumu ε_i s nulovou střední hodnotou a konstantním rozptylem. Časovou řadu tedy můžeme zapsat v tomto tvaru

$$y_t = \varepsilon_t + \psi_1 \varepsilon_{t-1} + \psi_2 \varepsilon_{t-2} + \dots \quad (2.19)$$

kde ψ_i jsou parametry.

Často se v tomto zápise využívá operátoru zpětného posunutí definovaného jako

$$B^j y_t = y_{t-j}. \quad (2.20)$$

Využitím tohoto operátoru lze lineární proces (2.19) přepsat do tvaru

$$y_t = \psi(B) \varepsilon_t, \quad (2.21)$$

kde $\psi(B) = 1 + \psi_1 B + \psi_2 B^2 + \dots = 1 + \sum_{j=1}^{\infty} \psi_j B^j$.

Lineární proces může existovat pouze tehdy, když nekonečná řada na pravé straně vztahu (2.21) konverguje. Z tohoto vyplývá, že zcela postačující podmínkou pro tuto existenci je, pokud $\psi(B)$ konverguje pro $|B| \leq 1$, viz [6].

Prakticky je dále vyžadováno, aby se pomocí minulých hodnot časové řady dala vyjádřit hodnota současná y_t . Tedy, aby

$$y_t = \pi_1 y_{t-1} + \pi_2 y_{t-2} + \dots + \varepsilon_t. \quad (2.22)$$

Výraz 2.22 lze vyjádřit opět pomocí operátoru zpětného posunutí, tedy v tomto tvaru

$$\pi(B)y_t = \varepsilon_t, \quad (2.23)$$

kde platí $\pi(B) = 1 - \pi_1 B - \pi_2 B^2 - \dots = 1 - \sum_{j=1}^{\infty} \pi_j B^j$. Lineární proces, který je možné vyjádřit

za použití minulých hodnot, nazýváme lineární proces *invertibilní*. Pokud z výrazu (2.23) osamostatníme y_t , tedy $y_t = \pi(B)^{-1} \varepsilon_t$ a porovnáme s výrazem 2.21, dostaneme následující rovnost

$$\pi(B)\psi(B) = 1 \quad (2.24)$$

V praxi se pracuje se speciálními lineárními procesy, které vzniknou uvažováním pouze konečného počtu parametrů v modelu (2.19). Nejčastěji používané modely v Box-Jenkinsově metodologii jsou modely MA (proces klouzavých součtů), AR (autoregresní proces) a model ARMA (model smíšený).

2.6.4 Proces klouzavých součtů MA

Proces klouzavých součtů se v literaturách značí jako MA(q), z anglického *Moving Average*, kde q značí řád modelu. Řád klouzavých součtů určuje počet parametrů v modelu.

Model klouzavých součtů má obecně tvar

$$y_t = \varepsilon_t + \vartheta_1 \varepsilon_{t-1} + \dots + \vartheta_q \varepsilon_{t-q}, \quad (2.25)$$

kde ε_t značí opět bílý šum a ϑ_j značí parametry tohoto modelu pro $j = 1, 2, \dots, q$. Model lze přepsat pomocí operátoru zpětného posunutí do tvaru

$$y_t = \vartheta_B \varepsilon_t, \quad (2.26)$$

kde $\vartheta_B = 1 + \sum_{j=1}^q \vartheta_j B^j$.

Pro další odvozování u modelů konkrétních řádů je nutné znát vzorec pro autokorelační funkci ρ_k u obecného procesu klouzavých součtů. Rozptyl tohoto procesu je jednoduché odvodit ve tvaru

$$\sigma_y^2 = \gamma_0 = (1 + \vartheta_1^2 + \dots + \vartheta_q^2) \sigma_\varepsilon^2. \quad (2.27)$$

Dále je zapotřebí odvodit kovarianci a to následujícím způsobem

$$\text{cov}(y_t, y_{t+k}) = (\vartheta_k + \vartheta_1 \vartheta_{k+1} + \dots + \vartheta_{q-k} \vartheta_q) \sigma_\varepsilon^2. \quad (2.28)$$

Po dosazení výrazů (2.27) a (2.28) do obecného vzorce pro výpočet autokorelační funkce (2.14) dostaneme následující výraz

$$\rho_k = \frac{\vartheta_k + \vartheta_1 \vartheta_{k+1} + \dots + \vartheta_{q-k} \vartheta_q}{1 + \vartheta_1^2 + \dots + \vartheta_q^2} \quad \text{pro } k = 1, 2, \dots, q \quad (2.29)$$

$$\rho_k = 0 \quad \text{pro } k > q$$

Je tedy zřejmé, že identifikačním bodem pro proces klouzavých součtů MA(q) je $k_0 = q$.

Proces MA(1)

Tento proces zapisujeme ve tvaru

$$y_t = \varepsilon_t + \vartheta_1 \varepsilon_{t-1}. \quad (2.30)$$

Proces lze zapsat i v invertibilním tvaru a to za podmínky, že $|\vartheta_1| < 1$, tedy

$$y_t = \vartheta_1 y_{t-1} - \vartheta_1^2 y_{t-2} + \vartheta_1^3 y_{t-3} - \dots + \varepsilon_t. \quad (2.31)$$

Autokorelační funkce po dosazení do (2.29) má tvar

$$\rho_1 = \frac{\vartheta_1}{1 + \vartheta_1^2}, \quad (2.32)$$

$$\rho_k = 0 \quad \text{pro } k > 1.$$

Je tedy zřejmé, že identifikačním bodem tohoto procesu $k_0 = 1$. Ze vztahu (2.32) lze odvodit podmínku $|\rho_1| < \frac{1}{2}$. Z rovnice (2.32) můžeme také odvodit vztah pro výpočet, resp. odhad parametru ϑ_1 procesu MA(1) pro případ, že do tohoto vztahu dosadíme odhadnutou hodnotu $\hat{\rho}_k = r_k$. Odhad parametru ϑ_1 tedy spočítáme dle vzorce

$$\vartheta_1 = \frac{1 - \sqrt{1 - 4r_1^2}}{2r_1}. \quad (2.33)$$

U procesu MA(2) bychom postupovali obdobně při stanovování jednotlivých charakteristik a omezení, proto se v této práci dalším odvozováním zabývat nebudeme.

2.6.5 Autoregresní proces AR(p)

Tento proces je definován předpisem

$$y_t = \zeta_1 y_{t-1} + \zeta_2 y_{t-2} + \dots + \zeta_p y_{t-p} + \varepsilon_t, \quad (2.34)$$

kde p značí řád modelu. Ze zápisu (2.34) je zřejmé, že autoregresní proces je již zapsán v invertibilním tvaru, proto se automaticky předpokládá jeho invertibilita. Tento proces je možné zapsat i pomocí operátoru zpětného posunutí takto

$$\xi(B)y_t = \varepsilon_t, \quad (2.35)$$

kde $\xi(B) = 1 - \sum_{j=1}^p \xi_j B^j$.

Při odvozování autokorelační funkce se vychází z diferenčních rovnic, které získáme těmito kroky:

- 1) Rovnici (2.34) vynásobíme výrazem y_{t-k} a dále budeme uvažovat přechod ke středním hodnotám
- 2) Víme, že střední hodnota bílého šumu je $E(\varepsilon_t) = 0$ a proto i $E(y_{t-k} \varepsilon_t) = 0$
- 3) Celou rovnici vydělíme výrazem $\gamma_0 = \text{cov}(y_t, y_{t+0})$

Nakonec získáme soustavu diferenčních rovnic ve tvaru

$$\rho_k = \xi_1 \rho_{k-1} + \xi_2 \rho_{k-2} + \dots + \xi_p \rho_{k-p}. \quad (2.36)$$

Pokud bychom dosazovali za $k = 1, 2, \dots, p$, získali bychom *Yuleovu-Walkerovu* soustavu rovnic, viz [8]. Řešením této soustavy získáme parametry $\xi_1, \xi_2, \dots, \xi_p$, vyjádřené pomocí hodnot autokorelační funkce.

Autoregresní proces AR(1)

Tento proces zapisujeme ve tvaru

$$y_t = \xi_1 y_{t-1} + \varepsilon_t. \quad (2.37)$$

Aby proces konvergoval, musí parametr ξ_1 splňovat podmínku $|\xi_1| < 1$. Pokud je tato podmínka splněna, můžeme jednoduše proces (2.37) přepsat do tvaru

$$y_t = \varepsilon_t + \xi_1 \varepsilon_{t-1} + \xi_1^2 \varepsilon_{t-2} + \dots \quad (2.38)$$

Dosažením za $k = 1$ do *Yuleovy-Walkerovy* soustavy rovnic dostaneme

$$\rho_1 = \xi_1 \rho_0. \quad (2.39)$$

Dále víme, že $\rho_0 = 1$ a proto

$$\rho_1 = \xi_1. \quad (2.40)$$

Stejným způsobem bychom odvodili charakteristiky a omezení pro proces AR(2), proto se tímto již nadále nebudeme zabývat.

2.6.6 Smíšený proces

Při modelování chování časových řad se v praxi setkáváme s případy, kdy stacionární náhodné procesy nelze zcela přesně popsat autoregresním modelem (AR) nebo modelem klouzavých součtů (MA). Tato situace je zřejmá z autokorelační (ACF) nebo parciálně autokorelační funkce (PACF), viz [9]. Z výše uvedeného vyplývá možné užití kombinace procesů AR a MA s dvojicí parametrů p, q . Model značíme ARMA(p, q) a označujeme jej jako *smíšený proces řádu p a q*. Proces je definován takto

$$y_t = \xi_1 y_{t-1} + \dots + \xi_p y_{t-p} + \varepsilon_t + \vartheta_1 \varepsilon_{t-1} + \dots + \vartheta_q \varepsilon_{t-q}. \quad (2.41)$$

Stejně jako u předešlých procesů, kdy bylo možné tyto procesy vyjádřit pomocí operátoru zpětného posunutí, je možné toto vyjádření také pro proces ARMA a to v následujícím tvaru

$$\xi(B)y_t = \vartheta(B)\varepsilon_t, \quad (2.42)$$

kde $\xi_{(B)}$ a $\vartheta_{(B)}$ jsou odvozeny na straně 36 a 38 této práce. Podmínka stacionarity smíšeného procesu $ARMA(p, q)$ je totožná s podmínkou stacionarity procesu $AR(p)$ a podmínka invertibility je totožná s podmínkou invertibility procesu $MA(q)$ jak uvádí Cipra v [6].

2.6.7 Fáze zpracování modelu

Celý proces tvorby modelu časové řady užitím Box-Jenkinsovy metodologie sestává ze 3 základních fází:

- 1) Identifikace modelu – v této fázi se rozhoduje o typu procesu a určení jeho řádu. Samotná identifikace spočívá již i v přípravných akcích pro zajištění stacionarity apod. V této fázi pracujeme s grafickým pojetím dané časové řady. Jak již bylo řečeno v textu výše, stacionaritu lze ověřit pouhým vizuálním ohledáním grafického znázornění dané časové řady. Pokud se nám podaří na základě studia autokorelačních a parciálních autokorelačních funkcí identifikovat model, můžeme dále provést počáteční odhady, viz tabulka 2.1.

Tabulka 2.1 - Odhady parametrů a omezující podmínky

PROCES	$MA(1)$	$MA(2)$	$AR(1)$	$AR(2)$
Odhad parametrů	$\hat{\vartheta}_1 = \frac{1 - \sqrt{1 - 4r_1^2}}{2r_1}$	Řeší se iteračně	$\hat{\xi}_1 = r_1$	$\hat{\xi}_1 = \frac{r_1(1 - r_2)}{1 - r_1^2}, \hat{\xi}_2 = \frac{r_2 - r_1^2}{1 - r_1^2}$
Omezení pro odhad r_k	$ r_1 < \frac{1}{2}$	$r_2 + r_1 > -\frac{1}{2}$ $r_2 - r_1 > -\frac{1}{2}$ $r_1^2 < 4r_2(1 - 2r_2)$	$ r_1 < 1$	$r_1^2 < \frac{r_2 + 1}{2}$ $ r_2 < 1$

Zdroj: Cipra [6]

- 2) Odhad parametrů – předešlé odhady zmíněné v tabulce 2.1 jsou pouze počátečními odhady, které je nutné dále upravovat např. metodou *podmíněných nejmenších*

čtverců případně *nepodmíněnou metodou nejmenších nelineárních čtverců*. Těmito metodami se nadále nebudeme zabývat, viz např. [6], [7], [8].

3) Ověřování modelu – tato fáze je také označována jako fáze diagnostická, jelikož má za účel analyzovat vhodnost navrženého modelu. Pokud by v této fázi byly nalezeny nesrovnalosti, bylo by zapotřebí všechny fáze zopakovat s jistými úpravami, které by právě na této fázi byly závislé. V této fázi je možné užít několik metod (převzato z [6]):

- Metoda přeparametrizování modelu- pokud by směrodatné odchylky odhadů parametrů byly neúměrně vysoké, je možné zvýšit počet parametrů, tedy přeparametrizovat stávající model
- Metoda odhadnutých reziduí- zkoumá chování autokorelační funkce odhadnuté reziduální složky a poměruje její hodnoty s dvojnásobkem směrodatné odchylky těchto odhadů – blíže je tato metoda popsána v [11]
- Portmanteau test - porovnává *portmanteau statistiku*, viz [6] a [7], tato statistika má chí-kvadrát rozdělení χ^2_{K-p-q} . Pokud hodnota statistiky překročí kritickou hodnotu $\chi^2_{K-p-q}(\alpha)$, pak na dané hladině významnosti α navržený model s konstatováním, že je není vhodný. Hodnota K odpovídá výrazu \sqrt{n} . Modifikací tohoto testu je Ljung-Boxův test, blíže popsany na str. 61 této práce.

Poslední dvě metody budou užity v praktické části, kde výsledky jednotlivých metod budou porovnány.

2.6.8 Předpovědi v Box-Jenkinsově metodologii

Je zřejmé, že přístup k modelování předpovědi bude pro Box-Jenkinse principiálně odlišný a při pochopení podstaty tohoto modelování jsou předpovědi relativně jednoduše konstruovatelné.

Označme $\hat{y}_{t+k}(t)$ předpověď pro hodnotu y_{t+k} , kterou konstruujeme v čase t , tj. předpověď v čase t o k kroků dopředu. V některých literaturách mluvíme o k jako o horizontu predikce. Obecně konstrukce předpovědí má lineární charakter, tedy hledáme takovou lineární funkci $y_t, y_{t-1}, y_{t-2}, \dots$, resp. $\varepsilon_t, \varepsilon_{t-1}$, která by minimalizovala střední čtvercovou chybu MSE, kde platí

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

Lineární funkci pro předpověď, kterou hledáme, můžeme zapsat takto

$$\hat{y}_{t+k}(t) = \psi_k^* \varepsilon_t + \psi_{k+1}^* \varepsilon_{t-1} + \dots \quad (2.43)$$

Pro obecný lineární proces můžeme psát

$$y_{t+k} = \varepsilon_{t+k} + \psi_1 \varepsilon_{t+k-1} + \psi_2 \varepsilon_{t+k-2} + \dots \quad (2.44)$$

Dosazením výrazů (2.43) a (2.44) do vzorce pro MSE, dostaneme následující účelovou funkci pro minimalizaci

$$(1 + \psi_1^2 + \psi_2^2 + \dots + \psi_{k-1}^2) \sigma_\varepsilon^2 + \sum_{i=0}^{\infty} (\psi_{t+i} - \psi_{t+i}^*)^2 \sigma_\varepsilon^2. \quad (2.45)$$

Abychom tento výraz minimalizovali, je zřejmě nutné položit $\psi_i = \psi_i^*$, na základě čehož lze dále psát

$$\hat{y}_{t+k}(t) = \psi_k \varepsilon_t + \psi_{k+1} \varepsilon_{t-1} + \dots \quad (2.46)$$

Dále lze psát [6]

$$\hat{y}_{t+k} = y_{t+k}, \quad (2.47)$$

kde lomené závorky značí podmíněné střední hodnoty při pevných hodnotách y_t, y_{t-1}, \dots

Označme nyní chybu předpovědi $e_{t+k}(t)$, jedná se o chybu předpovědi v čase t s horizontem predikce k . Můžeme tedy psát

$$e_{t+k}(t) = y_{t+k} - \hat{y}_{t+k}(t). \quad (2.48)$$

Je zřejmé, že rozptyl těchto chyb je

$$\text{var}\{e_{t+k}(t)\} = (1 + \psi_1^2 + \psi_2^2 + \dots + \psi_{k-1}^2) \sigma_\varepsilon^2 \quad (2.49)$$

a střední hodnota těchto chyb je nulová.

Uvažujme např. proces $ARMA(p,q)$, který je definován jako

$$y_t = \xi_1 y_{t-1} + \dots + \xi_p y_{t-p} + \varepsilon_t + \vartheta_1 \varepsilon_{t-1} + \dots + \vartheta_q \varepsilon_{t-q}. \quad (2.50)$$

O k kroků dopředu by model vypadal následovně

$$y_{t+k} = \xi_1 y_{t+k-1} + \dots + \xi_p y_{t+k-p} + \varepsilon_{t+k} + \vartheta_1 \varepsilon_{t+k-1} + \dots + \vartheta_q \varepsilon_{t+k-q}. \quad (2.51)$$

Budeme-li nyní hledat střední hodnotu tohoto modelu při pevně daných hodnotách y_t, y_{t-1}, \dots , dostaneme následující model

$$\hat{y}_{t+k}(t) = \xi_1 [y_{t+k-1}] + \dots + \xi_p [y_{t+k-p}] + [\varepsilon_{t+k}] + \vartheta_1 [y_{t+k-1}] + \dots + \vartheta_q [y_{t+k-q}]. \quad (2.52)$$

Prakticky je důležité uvědomit si několik souvislostí, na základě níž se skutečně konstruuje předpovědi (tyto souvislosti jsou dobře popsány v [6],[8],[12],[13]). Např., že očekávaná hodnota $y_{t+k} = y_{t+k}$ pro $k \leq 0$, tedy hodnotě skutečně odpozorované (využijeme tedy informací, které máme k dispozici z vypořezovaných hodnot) a $y_{t+k} = \hat{y}_{t+k}(t)$ pro $k > 0$, jedná se o hodnoty predikované. Dále

$[\varepsilon_{t+k}] = E[\varepsilon_{t+k}] = \varepsilon_{t+k} = y_{t+k} - \hat{y}_{t+k}(t+k-1)$, tedy chyba předpovědi pro $k \leq 0$ bude rovna rozdílu mezi napozorovanou a odhadnutou hodnotou, atd. Konstrukci předpovědi si názorně ukážeme na následujícím jednoduchém příkladě.

Uvažujme např. model ve tvaru $y_t = y_{t-1} + \varepsilon_t + 0,87\varepsilon_{t-1}$, tedy model ARMA (1,1).
 Předpověď o jeden krok dopředu by vypadala následovně

$$\hat{y}_{t+1}(t) = [y_t] + [\varepsilon_{t+1}] + 0,87 [\varepsilon_t].$$

Nyní je důležité položit ε_1 rovno nule [viz 6, str. 156], tak budeme moci začít rekurentní výpočet dosazováním za $t = 1, 2, \dots, n$. Předpovědi tedy vypadají následovně

$$\hat{y}_2(1) = y_1 + \varepsilon_2 + 0,87 \varepsilon_1 = y_1 + 0 + 0$$

$$\hat{y}_3(2) = y_2 + \varepsilon_3 + 0,87 \varepsilon_2 = y_2 + 0 + 0,87(y_2 - \hat{y}_2(1)).$$

Stejným způsobem bychom pokračovali dále. Pokud bychom měli k dispozici např. 150 dat, předpověď hodnoty 151 by vypadala následovně

$$\hat{y}_{151}(150) = y_{150} + \varepsilon_{151} + 0,87 \varepsilon_{150} = y_{150} + 0 + 0,87(y_{150} - \hat{y}_{150}(149)).$$

PRAKTICKÁ ČÁST

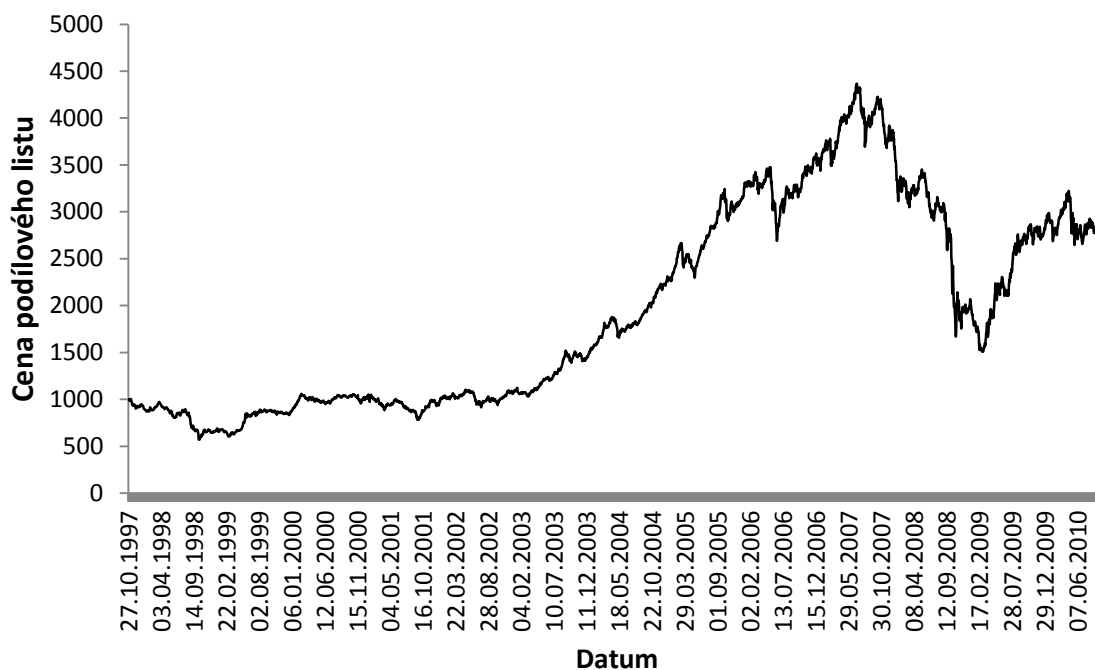
3 Analýza předpovědí

Pro analýzu předpovědí byly vybrány následující podílové fondy nabízené společností ING:

1) *ING International Český akciový fond (PF1)*

Jelikož se jedná o akciový fond, je tvořen nejméně z 51% akciemi, jenž jsou kótovány na BCPP (Burza cenných papírů Praha). Zbýlých, však nejvýše, 49% aktiv je tvořeno investicemi do akcií polských, maďarských, slovenských, rumunských, bulharských aj. Tento fond je převážně určen takovým investorům, jejichž cílem je dlouhodobé zhodnocení volných finančních prostředků (kapitálu) investovaných do právě těchto zmíněných evropských trhů. Vývoj tohoto fondu je zachycen na obr. 3.1.

Obrázek 3.1 – Vývoj ceny podílového listu (*ING International Český akciový fond*)

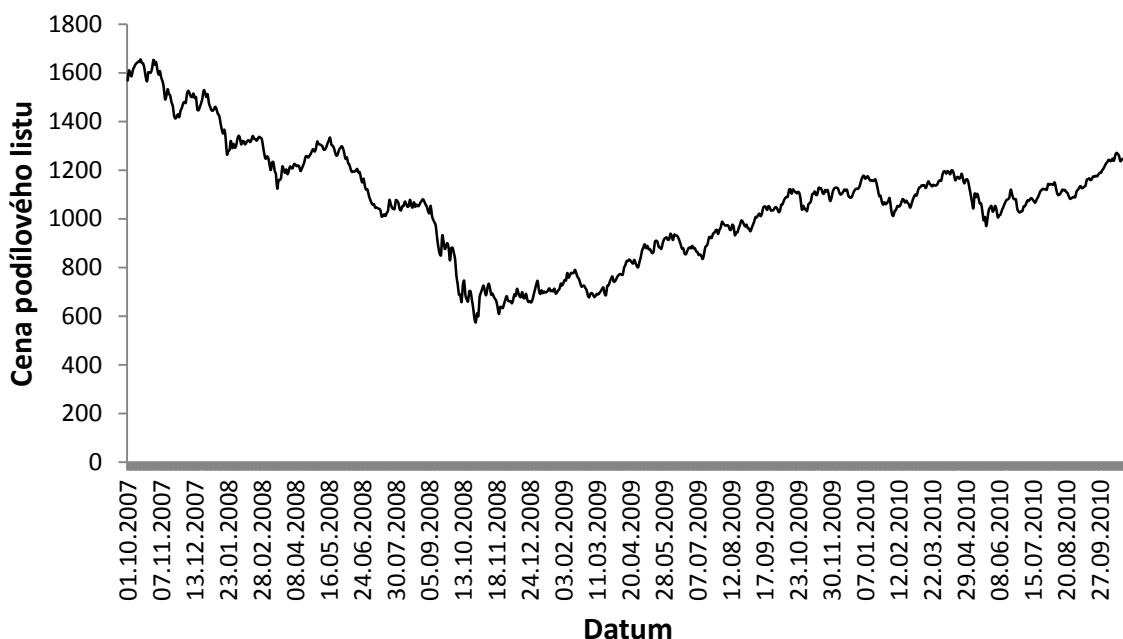


Zdroj: Vlastní zpracování dat získaných na www.ing.cz

2) *ING (L) Invest Emerging Markets (CZK) (PF2)*

Tento podílový fond je oproti prvním fondu směřován do zemí Latinské Ameriky, Asie, Středního Východu a Afriky. Jedná se tedy o geograficky diverzifikované portfolio. Důležitou informací o tomto fondu je fakt, že je tento fond zajištěn proti měnovému riziku. Vývoj tohoto fondu je zachycen na obr. 3.2.

Obrázek 3.2 – Vývoj ceny podílového listu (*ING (L) Invest Emerging Markets (CZK)*)



Zdroj: Vlastní zpracování dat získaných z www.ing.cz

U obou těchto fondů je zřejmé, že je k dispozici velice vysoké množství historických informací (dat) pro jejich analýzu. Jedná se totiž o prakticky jednodenní informace popisující cenu podílového listu. Takováto frekvence získávání dat je však pro účely této práce zbytečná. Vzhledem k tomu, že cílem analýzy má být stanovení predikcí budoucího vývoje a víme-li, že investování do podílových fondů je dlouhodobější záležitostí, není možné, ba ani účelné stanovovat predikce na horizont několika málo dnů. Proto je výhodnější data filtrovat, tzn. vybrat např. pouze měsíční či čtvrtletní data. Informace, která se tímto filtrováním vytratí, nemá viditelný vliv na výsledky analýzy. Proto jsem se při klasické dekompozici rozhodl využít pouze dvoutříměsíční data. Při užití

Box-Jenkinsovy metodologie jsou data měsíční (kvůli omezující podmínce na minimální počet pozorování).

3.1 Dekompoziční přístup pro PF1

Trendová složka

V první fázi je důležité identifikovat trendovou složku. Pro nalezení vhodné matematické křivky vystihující trend dané časové řady, bylo nutné počítat charakteristiky zmíněné v kapitole 2.4. Následující tabulka 3.1 shrnuje dosažené výsledky u vybraných trendových funkcí. Hodnoty původní časové řady jsou zachyceny v tabulce 3a, přílohy A.

Tabulka 3.1- Odhady trendových funkcí

	Odhad trendové funkce	Index determinace/ upravený index determinace	SSE	MSE
LINEÁRNÍ	$\hat{y}_t = 58,24t + 405,9$	0,66/0,63	19 761 532	380 029
KVADRATICKÝ	$\hat{y}_t = -0,547t^2 + 87,24t + 144,9$	0,68/0,67	19 130 275	367 889
EXPONENCIÁLNÍ	$\hat{y}_t = 678,7e^{0,033t}$	0,77/0,76	25 772 694	495 628
POLYNOM 3. STUPNĚ	$\hat{y}_t = -0,156t^3 + 11,92t^2 - 179,5t + 1379$	0,83/0,81	10 148 038	195154
POLYNOM 4. STUPNĚ	$\hat{y}_t = -0,00099t^4 - 0,059t^3 + 8,594t^2 - 139,512t + 1263$	0,83/0,81	10 096 210	194159

Zdroj: Vlastní zpracování dat

Z tabulky 3.1 je patrné, že nejvyšší index determinace (i upravený index determinace) a zároveň nejnižší hodnoty SSE a MSE byly dosaženy u polynomu 3. a 4. stupně, kdy hodnoty SSE a MSE jsou jen málo odlišné. Jelikož hodnoty predikcí u polynomu 4. stupně rychle klesají a již v polovině predikovaného období by dosáhly nuly, rozhodl jsem se pro užití polynomu 3. stupně. U tohoto polynomu také otestuji významnost jednotlivých parametrů. Půjde tedy o testování hypotézy $H_0: \beta_j = 0$ na hladině významnosti

$\alpha = 0,05$. Jinými slovy testujeme, zdali vysvětlující proměnná nemá vliv na vysvětlovanou proměnnou. Testovací statistika vypadá následovně

$$t_j = \frac{\hat{\beta}_j - \beta_j}{s_{b_j}}$$

kde $\hat{\beta}_j$ značí odhad parametru a s_{b_j} je směrodatná odchylka. Hodnoty testovací statistiky v absolutní hodnotě poté porovnááme s kritickou hodnotu studentova rozdělení s $n - (m + 1)$ stupni volnosti, kde m značí počet regresorů v modelu. Je-li hodnota testovací statistiky větší než kritická hodnota studentova rozdělení, nulovou hypotézu zamítáme na hladině významnosti $\alpha / 2$. Výsledky jsou zachyceny v následující tabulce.

Tabulka 3.2 – Testování významnosti parametrů polynomu 3. stupně

Parametr	Hodnota testovací statistiky	Kritická hodnota	Nulová hypotéza na hladině významnosti $\alpha = 0,05$
$\hat{\beta}_0$	2337,29	2,012	Zamítáme
$\hat{\beta}_1$	1795	2,012	Zamítáme
$\hat{\beta}_2$	8,96	2,012	Zamítáme
$\hat{\beta}_3$	0,096	2,012	Nezamítáme

Zdroj: Vlastní zpracování dat

Z tabulky 3.2 je zřejmé, že parametr $\hat{\beta}_3$ je nevýznamný, a je tedy možné jej z modelu vypustit. Poté je zapotřebí model přepočítat s vynecháním tohoto parametru. Hodnoty tedy odpovídají polynomu 2. stupně v tabulce 3.1. Tato trendová funkce bude použita k analýze periodické složky.

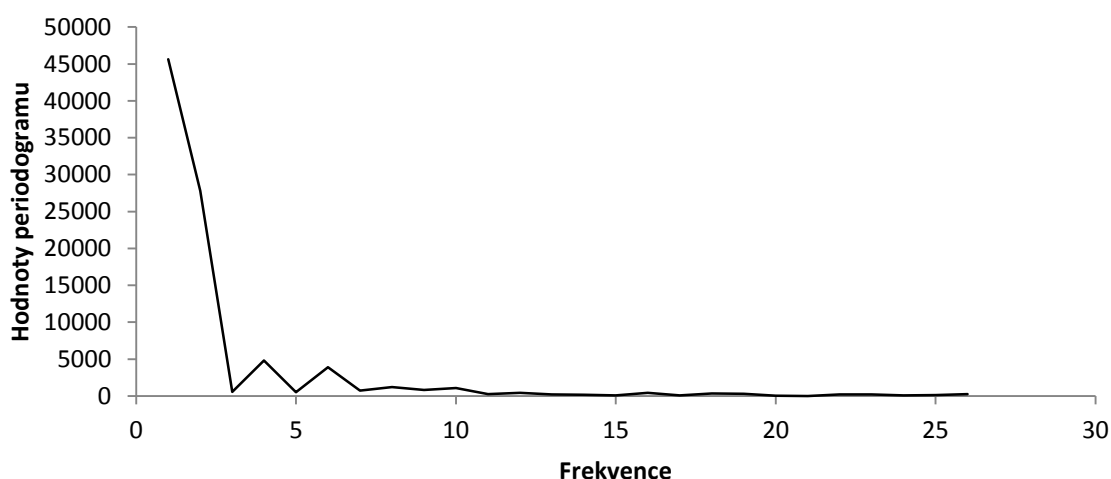
Periodická složka (model skrytých period)

Abychom mohli hledat významné periody v časové řadě pro PF1, musíme nejprve očistit tuto řadu od trendové složky (polynom 2. stupně). Dále tedy budeme pracovat s časovou řadou

$$y_t^{(o)} = y_t - \hat{y}_t.$$

Tuto očištěnou řadu použijeme pro výpočet hodnot periodogramu a následnou identifikaci významných period. Data pro zpracování podílového fondu 1 jsou zachycena v tabulce 3a, třetí sloupec. U této časové řady jsem se rozhodl pracovat s tříměsíčními daty získanými vždy na konci měsíce. Rozsah původní časové řady čítá cca 3500 dat, proto byla provedena až tříměsíční eliminace. Nyní je zapotřebí sestavit periodogram pro stanovení významných period pro zanesení do modelu. Tabulka 4a v příloze A shrnuje výpočty pro sestavení periodogramu. Nyní je možné zkonstruovat periodogram, viz obr. 3.3. Ze zpracovaného periodogramu lze vyčíst, že v dané řadě existuje několik Fourierových period. Zkusme tento předpoklad ověřit Fisherovým testem.

Obrázek 3.3 - Periodogram



Zdroj: Vlastní zpracování dat

Z tabulky 4a vyplývá, že nejvyšší hodnoty ν_j bylo dosaženo pro periodu s pořadovým číslem $j = 1$, a proto pomocí Fisherova testu budeme testovat nulovou hypotézu $H_0: y_t = a_0 + \varepsilon_t$ na hladině významnosti $\alpha = 0,05$. Kritické hodnoty pro Fisherův test jsou znázorněny v tabulce 2b, přílohy B. Hodnoty pro m , které nejsou uvedeny v této tabulce lze získat aproximací. Kritická hodnota tedy je $g_F(0,05) = 0,216$, jelikož $m=26$. Hodnota testovací statistiky je

$$F = \frac{\nu_1}{\sum_{j=1}^{26} \nu_j} = \frac{45644,80}{90193,65} = 0,506$$

Porovnáme-li hodnotu statistiky s kritickou hodnotou, vidíme, že $F > g_F(0,05)$ a tudíž nulovou hypotézu zamítáme na hladině významnosti 0,05. Objevili jsme tedy první významnou periodu v modelu. Nyní otestujeme druhou největší hodnotu periodogramu. Statistika vypadá následovně

$$F = \frac{v_2}{\sum_{j=2}^{26} v_j} = \frac{27811,71}{90193,65 - 45644,80} = 0,624$$

a porovnáme-li hodnotu statistiky s kritickou hodnotou, vidíme, že $F > g_F(0,05)$, jelikož $g_F(0,05) = 0,228$. Nulovou hypotézu o nevýznamnosti periody zamítáme opět na hladině významnosti 0,05. Pokračujeme v testování třetí největší hodnoty periodogramu, pro kterou vypadá statistika následovně

$$F = \frac{v_4}{\sum_{j=3}^{26} v_j} = \frac{4797,00}{90193,65 - 45644,80 - 27811,71} = 0,286.$$

Kritická hodnota byla aproximací stanovena jako $g_F(0,05) = 0,2341$. Platí tedy $F > g_F(0,05)$ a nulovou hypotézu tedy opět zamítáme s tím, že jsme našli 3. významnou periodu. Dále je zapotřebí stanovit hodnotu statistiky pro 4. největší hodnotu periodogramu, která vypadá následovně

$$F = \frac{v_6}{\sum_{\substack{j=3 \\ j \neq 4}}^{26} v_j} = \frac{3901,07}{90193,65 - 45644,80 - 27811,71 - 4797} = 0,326.$$

Kritická hodnota je $g_F(0,05) = 0,2532$, tedy nulovou hypotézu na hladině významnosti 0,05 zamítáme. Pokračujeme v testování 5. nejvyšší hodnoty periodogramu

$$F = \frac{v_8}{\sum_{\substack{j=3 \\ j \neq 4 \\ j \neq 6}}^{26} v_j} = \frac{1205,69}{90193,65 - 45644,80 - 27811,71 - 4797 - 3901,07} = 0,149.$$

Nyní je zřejmé, že nulovou hypotézu na hladině $\alpha = 0,05$ nezamítáme, jelikož kritická hodnota je $g_F(0,05) = 0,262$ a platí tedy, že $F < g_F(0,05)$. Proto celou proceduru ukončíme s tvrzením, že jsme našli 4 významné periody, které můžeme zanést do modelu.

Nalezli jsme tedy trendovou i periodickou složku a je tedy zapotřebí zkonstruovat predikční intervaly. Abychom však byli schopni zkonstruovat kvalitní predikce a predikční intervaly je zapotřebí znovu odhadnout kompletní model zahrnující jak trendovou, tak periodickou složku. Výsledné parametry pro model obsahující trendovou i periodickou složku jsou zachyceny v tabulce 3.3.

Tabulka 3.3 – Parametry modelu a jejich přesnosti

Parametr	β_0	β_1	β_2	a_1	b_1	a_2	b_2
Odhad	3539,65	-238,74	5,11	-923,86	-1476,69	233,37	-641,48
Přesnost	$\pm 2,874$	$\pm 0,313$	$\pm 0,005$	$\pm 0,420$	$\pm 1,625$	$\pm 0,269$	$\pm 0,446$
Parametr	a_4	b_4	a_6	b_6			
Odhad	-24,02	150,35	-155,87	-70,85			
Přesnost	$\pm 0,216$	$\pm 0,219$	$\pm 0,204$	$\pm 0,201$			

Zdroj: Vlastní zpracování dat

Nově zkonstruovaný model tedy čítá 11 odhadovaných parametrů a vypadá následovně

$$\hat{y}_t = 3539,65 - 238,74t + 5,11t^2 - 923,86 \sin(0,12t) - 1476,69 \cos(0,12t) + 233,37 \sin(0,24t) - 641,48 \cos(0,24t) - 24,02 \sin(0,48t) + 150,35 \cos(0,48t) - 155,87 \sin(0,72t) - 70,85 \cos(0,72t)$$

Zápis celkového modelu neobsahuje přesnosti, jelikož by tento model byl až příliš nepřehledný. Na základě tohoto nového modelu bylo možné spočítat bodové predikce a predikční intervaly, jež jsou uvedeny v tabulce 6a, příloha A. Jak je vidět z obr. 3.4, intervalová predikce je relativně úzká. Tento fakt přisuzuji velkému počtu nalezených významných period, které byly zaneseny do modelu, a dalo by se předpokládat, že daný model bude v predikování velice přesný. Sledovaná veličina se však do těchto intervalů nemusí vůbec vejít.

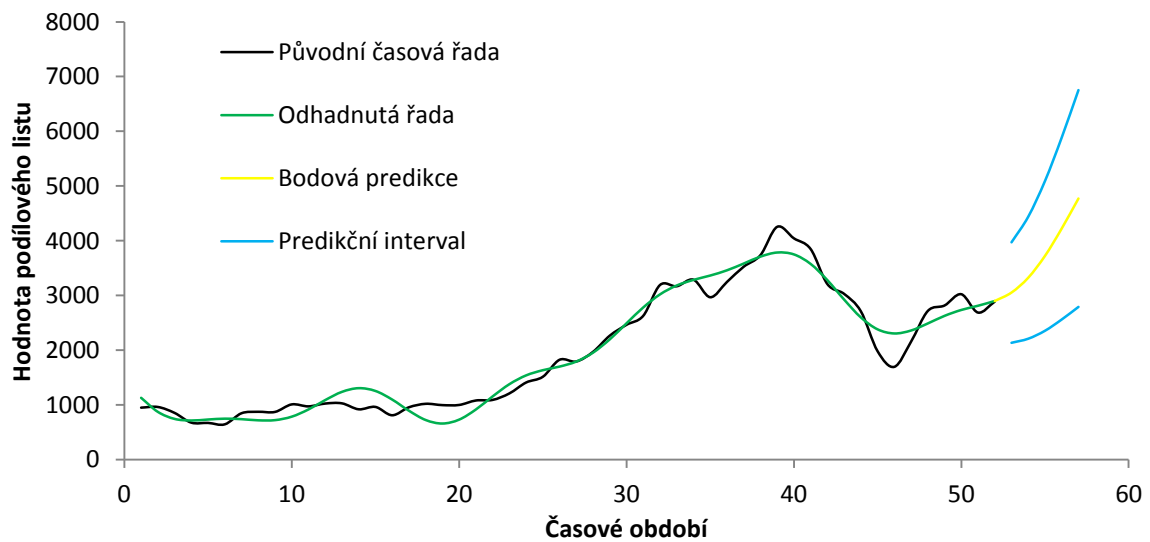
Dále je nutné ověřit předpoklad normality reziduí. V reálných situacích může být tento předpoklad porušen výskytem odlehlých pozorování, které tak znehodnocují výsledky získané z regresní analýzy. Pro testování normality zvolím Shapiro-Wilk test

studentizovaných reziduí. Tento test spočívá v uspořádání těchto reziduí $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$ vzestupně tak, že $\varepsilon_1 \leq \varepsilon_2 \leq \dots \leq \varepsilon_n$ a následném porovnání distribuční funkce těchto reziduí s teoretickou distribuční funkcí normálního rozdělení. Počítáme statistiku (převzato z [16])

$$W = \frac{\left[\sum_{i=1}^k a_i (\varepsilon_{n-i+1} - \varepsilon_i) \right]^2}{\sum_{i=1}^n (\varepsilon_i - \bar{\varepsilon})^2},$$

kde a_i jsou tabelované hodnoty vah, $k = (n + 1) / 2$ pro n liché a $k = n / 2$ pro n liché. Testujeme nulovou hypotézu $H_0: X \square N$. Pokud je hodnota statistiky W blízká 1 a zároveň p-value není značně nízká, jedná se o shodu distribučních funkcí a nulovou hypotézu na hladině významnosti $\alpha = 0,05$ nelze zamítnout. Hodnota statistiky pro PF1 má hodnotu $w = 0,9833$, přičemž $p(0,05) = 0,9470$. Nulovou hypotézu tedy nelze zamítnout na hladině významnosti $\alpha = 0,05$.

Obrázek 3.4 – Kompletní model s bodovou predikcí a predikčními intervaly



Zdroj: Vlastní zpracování dat

Index determinace pro tento výsledný model vypadá následovně

$$I^2 = \frac{\sum_{t=1}^n (\hat{y}_t - \bar{y}_t)^2}{\sum_{t=1}^n (y_t - \bar{y}_t)^2} = \frac{255888219,09}{257230406,10} = 0,99.$$

Je tedy zřejmé, že navržený model vystihuje danou časovou řadu výborně. Hodnota MSE je

$$MSE = \frac{\sum_{t=1}^n (y_t - \hat{y}_t)^2}{n} = 45471,46.$$

3.2 Dekompoziční přístup pro PF2

Trendová složka

Byl proveden výběr většího množství matematických křivek, které by eventuelně dobře popisovaly vývoj této časové řady. Jejich charakteristiky jsou zachyceny v tabulce 3.4. Charakteristika výběru spočívá v nalezení takové trendové funkce, která bude disponovat nejvyšším indexem determinace (upraveným indexem determinace) a nejnižší hodnotou SSE a tedy i MSE.

Tabulka 3.4 - Odhady trendové funkce

	Odhad trendové funkce	Index determinace/ upravený index determinace	SSE	MSE
LINEÁRNÍ	$\hat{y}_t = 1217,62 - 13,98t$	0,093/0,039	1 076 362	56650
KVADRATICKÝ	$\hat{y}_t = 7,51t^2 - 164,10t + 1743,31$	0,889/0,875	311 267	16382
EXPONENCIÁLNÍ	$\hat{y}_t = 1152,45e^{0,99t}$	0,059/0,036	1 090 903	57415
POLYNOM 3. STUPNĚ	$\hat{y}_t = -0,63t^3 + 26,43t^2 - 319,4t + 2034,01$	0,84/0,81	189 186	9957
POLYNOM 4. STUPNĚ	$\hat{y}_t = -0,06t^4 + 1,72t^3 - 4,30t^2 - 174,40t + 1856,13$	0,88/0,85	172 779	9093

Zdroj: Vlastní zpracování dat získaných z www.ing.cz

Jak je vidět z tabulky 3.4, těmto kritériím vyhovuje nejlépe polynom 4. stupně. Při analýze průběhu této funkce na intervalu $h = \langle T_{20}, T_{24} \rangle$, tj. na úrovni predikce jsem však zjistil, že tuto trendovou funkci není vhodné zvolit, neboť již v druhé polovině tohoto intervalu nabývá i záporné hodnoty (stejně jako u PF1). Cena podílového listu nemůže nabývat záporných hodnot. Proto volím druhou nejlépe vyhovující trendovou funkci -

polynom 3. stupně. Dále otestuji významnost jednotlivých parametrů stejným způsobem, jako u PF1. Výsledky testování jsou zachyceny v tabulce 3.5.

Tabulka 3.5 – Testování významnosti jednotlivých parametrů

Parametr	Hodnota testovací statistiky	Kritická hodnota	Nulová hypotéza
$\hat{\beta}_0$	1800,66	2,131	Zamítáme
$\hat{\beta}_1$	665,60	2,131	Zamítáme
$\hat{\beta}_2$	528,6	2,131	Zamítáme
$\hat{\beta}_3$	315	2,131	Zamítáme

Zdroj: *Vlastní zpracování dat*

Byla tedy potvrzena významnost jednotlivých parametrů.

Z tabulky 10a v příloze A vyplývá, pro která budoucí období chceme danou časovou řadu predikovat. Tabulka také obsahuje pořadová čísla 20, 21, 22, 23, 24, 25, 26, přiřazená jednotlivým datumům, které odpovídají vždy posledním dnům v daném měsíci, během kterých se obchodovalo na burzách, a během kterých se tedy vytvářela hodnota podílového listu.

Periodická složka (model skrytých period pro PF2)

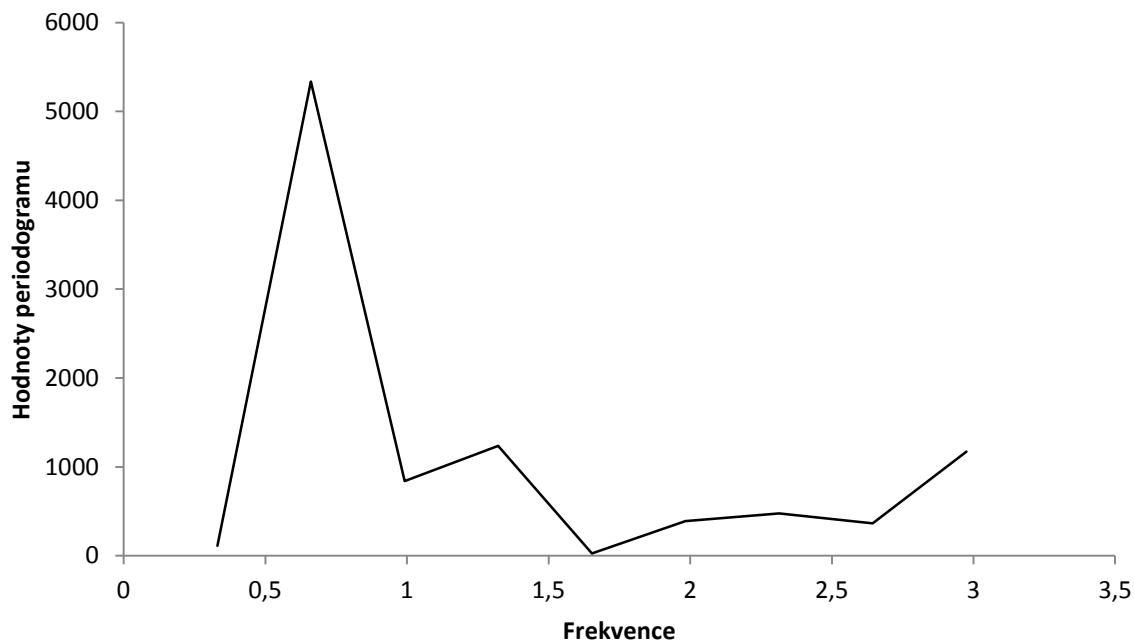
Stejně jako u PF1 je nutné nejdříve časovou řadu očistit od trendové složky. Výsledky tohoto očištění jsou znázorněny v tabulce 7a, přílohy A. Na základě této očištěné řady lze spočítat jednotlivé parametry pro výpočet hodnot periodogramu. Výpočty těchto hodnot jsou zachyceny v tabulce 8a v příloze A.

Z periodogramu (viz obr. 3.5) je zřejmé, že existuje alespoň jedna hodnota, která je výrazně odlišná od ostatních, proto je vysoká pravděpodobnost nalezení významné periody. Opět tedy užijeme Fisherův test a to nejprve pro $j = 2$. Tedy spočítáme testovací statistiku

$$F = \frac{v_2}{\sum_{j=1}^9 v_j} = \frac{5334,54}{9956,09} = 0,536.$$

Pokud hodnotu testovací statistiky porovnáme s aproximovanou kritickou hodnotou pro Fisherův test, dostaneme, že $0,536 > 0,469$, tedy $F > g_F(0,05)$ a nulovou hypotézu na hladině významnosti 0,05 zamítáme. Je tedy zřejmé, že perioda s pořadovým číslem $j = 2$ bude zahrnuta v modelu, nyní je zapotřebí otestovat i zbylé hodnoty v periodogramu.

Obrázek 3.5 - Periodogram



Zdroj: Vlastní zpracování dat

Otestujeme tedy druhou nejvyšší hodnotu periodogramu (hodnotu s pořadovým číslem $j = 4$). Testovací statistika je následující

$$F = \frac{v_4}{\sum_{\substack{j=1 \\ j \neq 2}}^9 v_j} = \frac{1237,29}{9956,09 - 5334,54} = 0,268.$$

Hodnotu testovací statistiky nyní porovnáme s $g_F = 0,5406$ a tedy platí $F < g_F(0,05)$. Nulovou hypotézu o nevýznamnosti dané periody na hladině významnosti $\alpha = 0,05$ nezamítáme. Nyní lze proceduru ukončit s tvrzením, že jsme našli jednu významnou periodu a to s pořadovým číslem $j = 2$, kterou zaneseme do modelu skrytých period. K tomu, abychom mohli spočítat kvalitní predikce (resp. predikční intervaly), bude nutné

odhadnout celkový model obsahující obě složky (trendovou a periodickou). Parametry spolu s jejich přesností jsou zachyceny v tabulce 3.6.

Tabulka 3.6 – Odhady parametrů a jejich přesnost

Parametr	β_0	β_1	β_2	β_3	a_2	b_2
Odhad	2030,28	-322,94	26,43	-0,62	-3,68	14,85
Přesnost	$\pm 1,13$	$\pm 0,48$	$\pm 0,05$	$\pm 0,002$	$\pm 0,33$	$\pm 0,33$

Zdroj: Vlastní zpracování dat

Nový odhadnutý model s trendovou i periodickou složkou tedy vypadá následovně

$$\hat{y}_t = 2030,28 - 322,94t + 26,43t^2 - 0,62t^3 - 3,68 \sin(0,66t) + 14,85 \cos(0,66t).$$

Přesnost jednotlivých parametrů, která by měla být zanesena do celkového modelu je uvedena v tabulce 3.6 a to z důvodu přehlednosti celkového modelu. Tento model je znázorněn na obr. 3.6 spolu s bodovou predikcí a predikčními intervaly. Předpovědi tohoto modelu byly zpracovány dle vztahu 2.11 a výsledky jsou zachyceny v tabulce 10a v příloze A. Koeficient determinace je pro tento model 0,83, což je poměrně vysoká hodnota, tudíž lze předpokládat, že tento model vystihuje charakter dané časové řady poměrně kvalitně. Negativním se zde jeví fakt, že šířka predikčního intervalu roste relativně rychle v závislosti na T, tedy na časových obdobích pro predikci. Tuto slabou stránku přisuzují pouze jediné nalezené významné periodě.

Dále je tedy zapotřebí otestovat normalitu studentizovaných reziduí jako u PF1. Hodnota testovací statistiky je $w = 0,9290$ a hodnota $p(0,05) = 0,9010$ je relativně vysoká. Nulovou hypotézu tedy nelze zamítnout na hladině významnosti $\alpha = 0,05$ a můžeme tedy konstatovat, že rezidua pochází z normálního rozdělení.

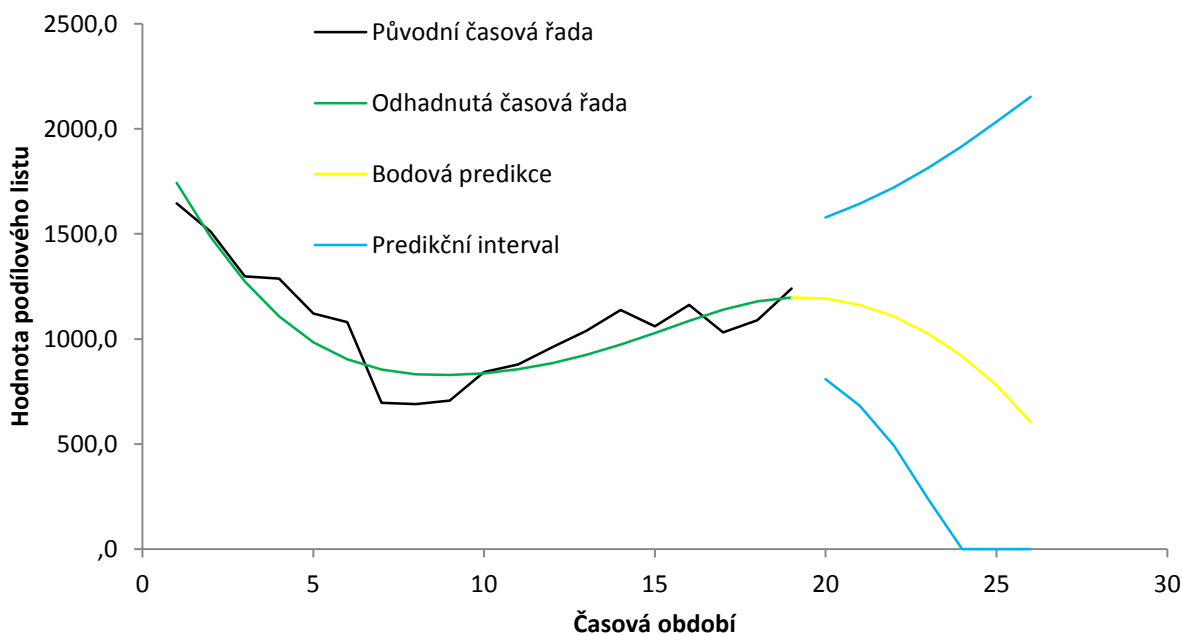
Index determinace pro tento model vypadá následovně

$$I^2 = \frac{\sum_{t=1}^n (\hat{y}_t - \bar{y}_t)^2}{\sum_{t=1}^n (y_t - \bar{y}_t)^2} = \frac{19765786,73}{20875505,90} = 0,95.$$

Opět vidíme, že daný model vystihuje chování analyzované řady velice dobře. Hodnota MSE je

$$MSE = \frac{\sum_{t=1}^n (y_t - \hat{y}_t)^2}{n} = 12059,53.$$

Obrázek 3.6 – Kompletní model s bodovou predikcí a predikčními intervaly



Zdroj: Vlastní zpracování dat

3.3 Box-Jenkinsova metodologie pro PF1

Při zpracování této metodologie byly uvažovány údaje za posledních 6 let. Byly vybrány hodnoty vždy k poslednímu obchodnímu dni v měsíci, viz tabulka 11 a příloha A. Při aplikaci této metodologie bylo nutné provést jiný výběr dat pro zpracování. Kdybych uvažoval dvouměsíční data jako u předchozích přístupů, nebyl by rozsah dat dostatečný pro aplikaci Box-Jenkinse, bylo tedy nutné selektovat data měsíční. Ke zpracování Box-

Jenkinsovy metodologie jsem použil software R, jelikož umožňuje velice rychlé zpracování dat s dobrou možností editace jednotlivých příkazů.

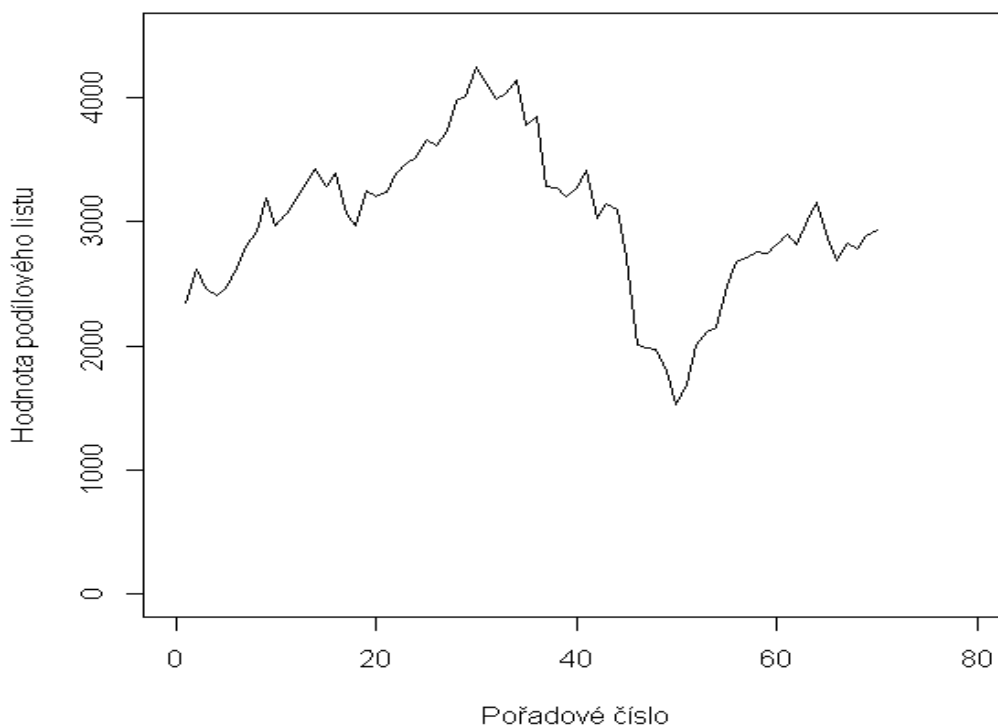
3.3.1 Ověřování stacionarity a identifikace modelu

Pro zobrazení selektovaného souboru dat byl použit software R za použití následujícího programového kódu

```
> PF1<- read.csv(„D:/PF1.csv“, header=T, dec=„,“, sep=„;“)
> plot(PF1[,1],type=„l“,xlim=c(0,80),ylim=c(0,4500)).
```

Grafické vyobrazení daného souboru je uvedeno na obr. 3.7. Jak je zřejmé z tohoto obrázku, tato řada není stacionární, a proto je nutné provést její transformaci diferencováním.

Obrázek 3.7 – Průběh ceny podílového listu (vybraná data)



Zdroj: Vlastní zpracování dat v R

Abychom potvrdili nestacionaritu původní časové řady a následnou stacionaritu transformované časové řady, je zapotřebí získat hodnoty autokorelační a parciální autokorelační funkce. V software R bylo zapotřebí vytvořit následující programový kód k otestování stacionarity

```
> par(mfrow=c(2,2))
> acf(PF1[,1],main=„Průběh ACF před diferencováním“,xlab=„k“,ylab=„rk“)
> acf(diff(PF1[,1]),main=„Průběh ACF po diferencování“,xlab=„k“,ylab=„rk“)
> pacf(PF1[,1],main=„Průběh PACF před diferencováním“,xlab=„k“,ylab=„rkk“)
> pacf(diff(PF1[,1]),main=„Průběh PACF po diferencování“,xlab=„k“,ylab=„rkk“).
```

Na obr. 3.8 jsou zachyceny průběhy autokorelační funkce časové řady před diferencováním a po diferencování (první řádek) a průběh parciální autokorelační funkce před a po diferencování (druhý řádek). Jelikož ani ACF, ani PACF nemají identifikační bod (hodnoty ACF a PACF jsou vždy menší než příslušná hranice 2σ daná přerušovanými modrými hranicemi na obr. 3.8), lze tento model identifikovat jako model $ARMA(p,q)$. Budeme tedy aplikovat model $ARMA(1,1)$ bez dalšího zkoumání.

3.3.2 Odhad parametrů modelu

K odhadu parametrů $\hat{g}_1, \hat{\xi}_1$ modelu $ARMA(1,1)$ jsem použil opět software R s využitím tohoto programového kódu

```
> PF1.fit<-arima(PF1[,1],order=c(1,0,1))
> PF1.fit.
```

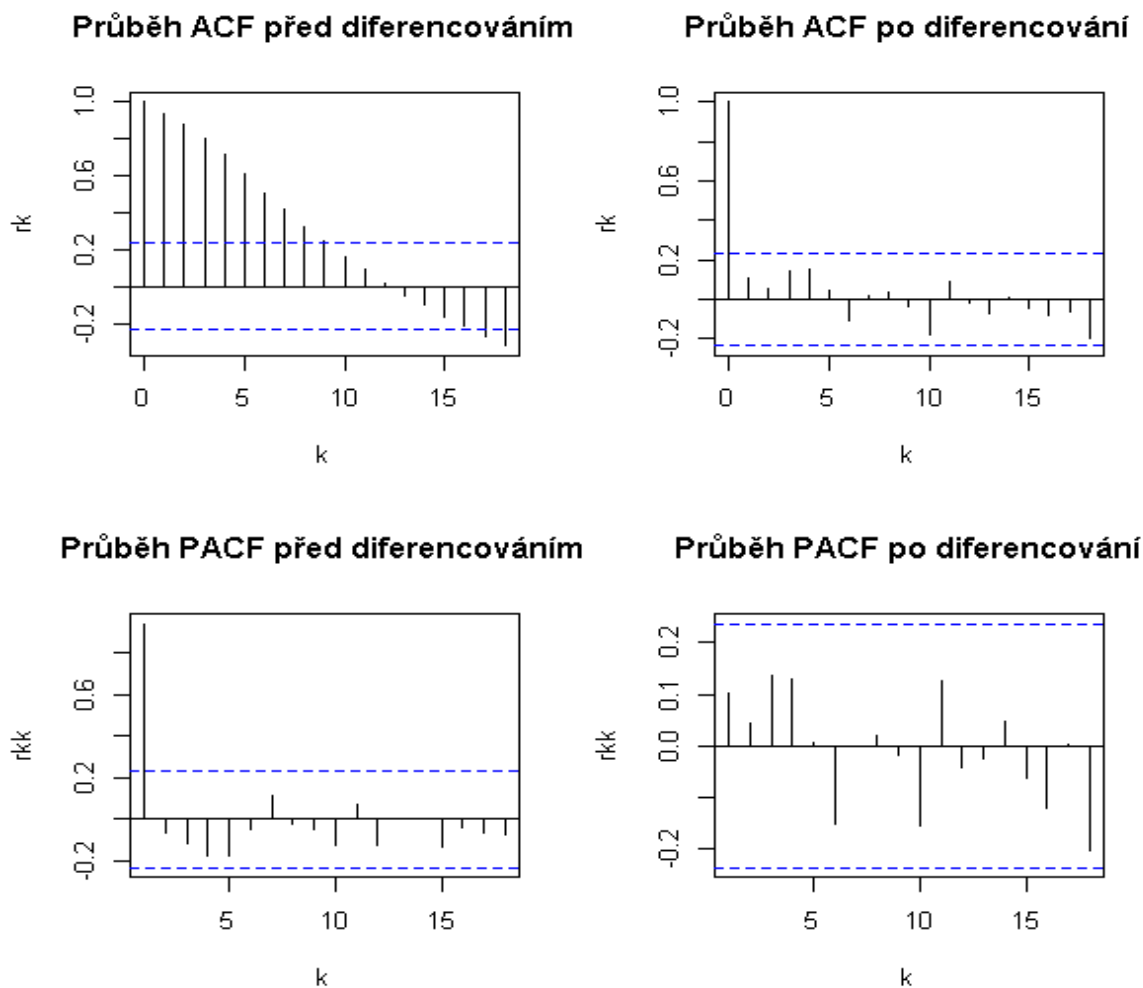
Spodní řádek je příkaz pro vrácení hodnot odhadů parametrů spolu s jejich směrodatnými odchylkami, tyto hodnoty jsou shrnuty v tabulce 3.7.

Tabulka 3.7 - Odhady parametrů

Parametr	$\hat{\xi}_1$	\hat{g}_1
Odhad	0,9316	0,1233
Přesnost	$\pm 0,0408$	$\pm 0,0219$

Zdroj: Vlastní zpracování dat v R

Obrázek 3.8 – Autokorelační a parciální autokorelační funkce



Zdroj: Vlastní zpracování dat v R

3.3.3 Diagnostika modelu

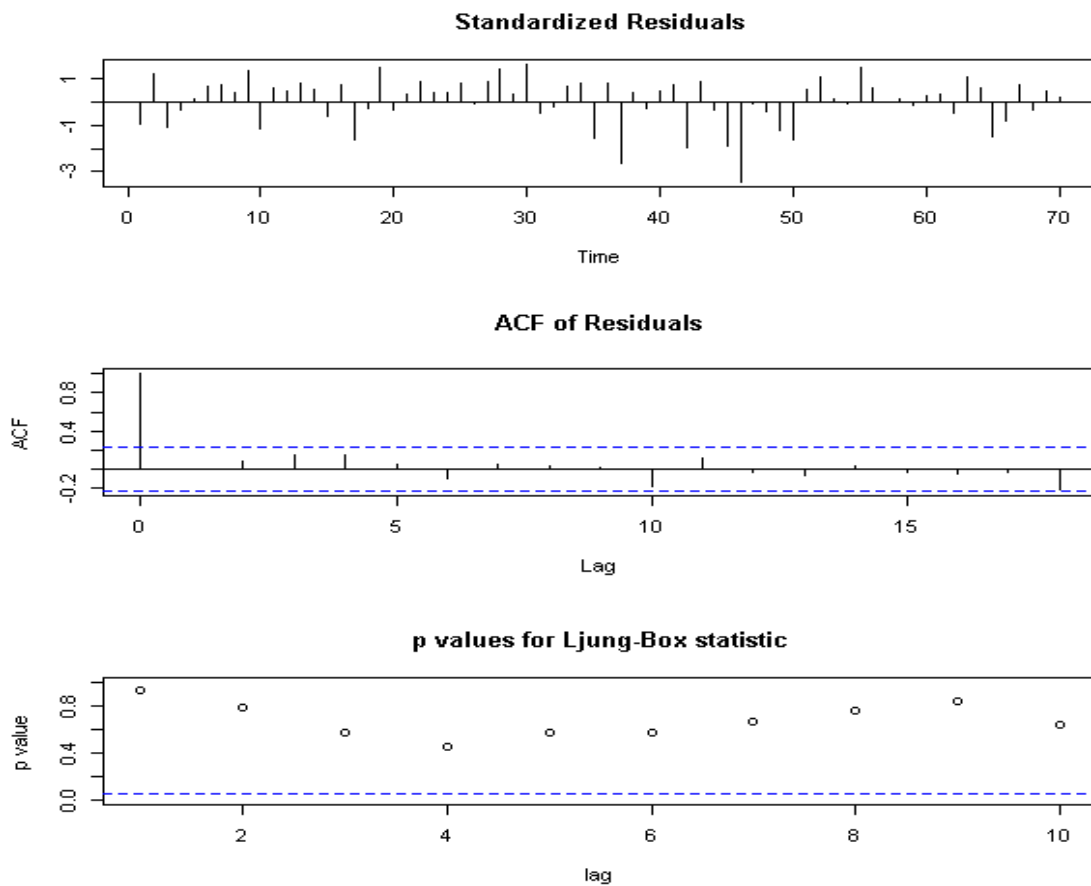
Diagnostiku modelu jsem se rozhodl zpracovat pomocí metody odhadnutých reziduí. Tato metoda spočívá v porovnávání hodnot autokorelační funkce odhadnuté reziduální složky (označme $\hat{r}_k(\mathcal{E})$) s dvojnásobkem jejich směrodatné odchylky. Pokud by, v absolutním vyjádření, některá z hodnot $\hat{r}_k(\mathcal{E})$ překročila dvojnásobek směrodatné odchylky, s velkou pravděpodobností je náš model nevyhovující.

Tuto diagnostiku lze také zpracovat v R pomocí následujícího programového kódu

```
> fit<-arima(PF1[,1],order=c(1,0,1))
> tsdiag(fit).
```

Spodní řádek je příkaz pro vykreslení odhadu reziduí a příslušné autokorelační funkce spolu se zakreslenými mezemi (dvojnásobky směrodatné odchyly). Z obr. 3.9 je vidět, že model $ARMA(1,1)$ je vyhovující, jelikož žádná z absolutních hodnot $r_k(\hat{\varepsilon})$ nepřekročila dvojnásobek směrodatné odchyly těchto reziduí.

Obrázek 3.9 – Diagnostika modelu $ARMA(1,1)$



Zdroj: Vlastní zpracování dat v R

Proto výsledný model pro podílový fond 1 vypadá následovně

$$y_t = (0,9316 \pm 0,0408)y_{t-1} + \varepsilon_t + (0,1233 \pm 0,0219)\varepsilon_{t-1}$$

Pro ujištění vyhodnotíme také Ljung-Boxovu statistiku. Tato statistika (zpracovaná ve spodní části obr. 3.9) je modifikací Portmanteau testu a testuje tedy autokorelaci reziduí. Statistiku počítáme dle následujícího vzorce (převzato z [13])

$$Q(m) = n \cdot (n + 2) \sum_{k=1}^m \frac{r_k^2}{n - k},$$

kde n značí rozsah souboru dat a r_k značí odhad koeficientu autokorelace se zpožděním k . Statistika má rozdělení χ_m^2 , tedy chí-kvadrát rozdělení s m stupni volnosti. Testujeme nulovou hypotézu $H_0: \rho_1 = \rho_2 = \dots = \rho_m = 0$ oproti alternativní hypotéze $H_1: \rho_k \neq 0$. Pro danou statistiku počítáme *p-value*, tedy pravděpodobnost, s jakou testovací statistika nabývá hodnot horších, než jaká je skutečná hodnota statistiky. Pokud *p-value* dosáhne hodnoty menší nebo rovné α , kde α značí hladinu významnosti, nulovou hypotézu na této hladině významnosti zamítáme. Jak je vidět z obr. 3.9, žádná z *p-values* není menší než hladina významnosti $\alpha = 0,05$ vyznačená modrou přerušovanou hranicí v témže obrázku (spodní část), a proto nulovou hypotézu nelze zamítnout. Model je tedy opravdu vyhovující.

3.3.4 Bodová a intervalová predikce

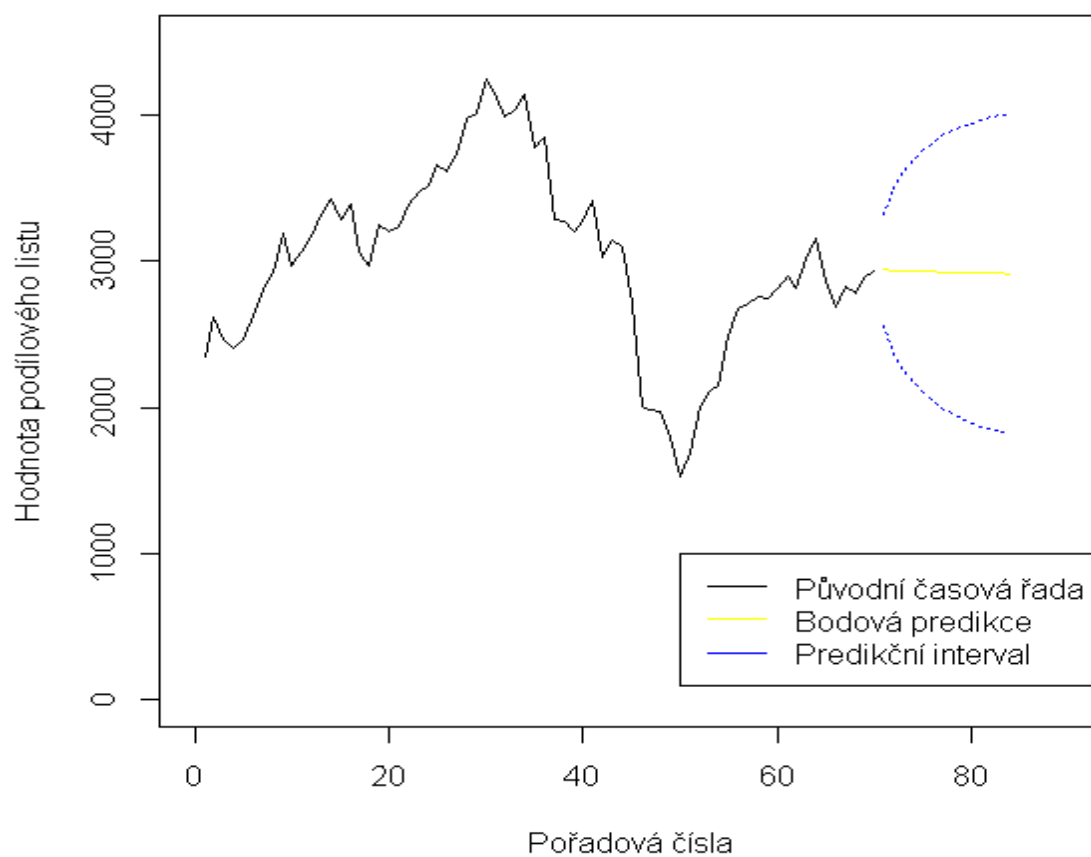
Bodová i intervalová predikce byla získána také programově v R za užití následujícího programového kódu (predikce jsou zobrazeny na obr. 3.10)

```
> plot(PF1[,1],type="l",xlim=c(0,80),ylim=c(0,4500),xlab="Pořadová čísla", ylab="Hodnota
podílového listu",main="Bodová+intervalová predikce")
> PF1.pred<-predict(PF1.fit,n.ahead=14)
> lines(PF1.pred$pred,col="yellow")
> lines(PF1.pred$pred-2*PF1.pred$se,col="blue",lty=3)
> lines(PF1.pred$pred+2*PF1.pred$se,col="blue",lty=3).
```

Konkrétní hodnoty bodové i intervalové předpovědi mohou být zpětně vyvolány pomocí následujících, po sobě jdoucích příkazů

```
> PF1.pred$pred  
> PF1.pred$pred+2*PF1.pred$se  
> PF1.pred$pred-2*PF1.pred$se.
```

Obrázek 3.10 – Bodová a intervalová predikce



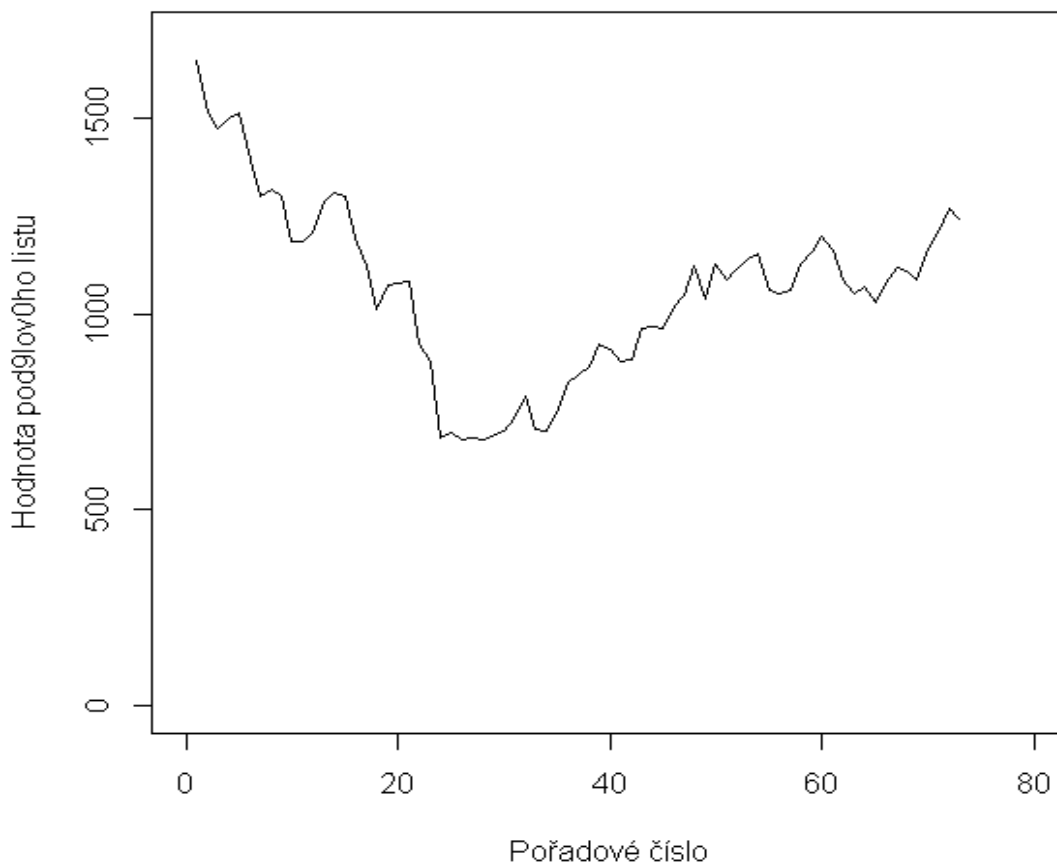
Zdroj: Vlastní zpracování dat v R

Číselné hodnoty bodových a intervalových předpovědí jsou zachyceny v tabulce 12a v příloze A.

3.4 Box-Jenkinsova metodologie PF2

U podílového fondu č. 2 bylo komplikovanější udělat selekci dat, která by měla být zahrnuta do modelování. Pouze měsíční data by opět tvořily neúplný soubor ($n < 50$), a tudíž by tato metodologie nemohla být použita. Rozhodl jsem se tedy pracovat s 15-ti denními pozorováními. Tato data jsou zachycena v tabulce 13a v příloze A. Je zřejmé, že z tohoto důvodu budu muset predikovat na mnoho období dopředu, tudíž se bude o to více snižovat kvalita předpovědí na takto dlouhá predikční období. Časová řada je zachycena na obr. 3.11.

Obrázek 3.11 – Vývoj ceny podílového listu



Zdroj: Vlastní zpracování dat v R

3.4.1 Ověřování stacionarity a identifikace modelu

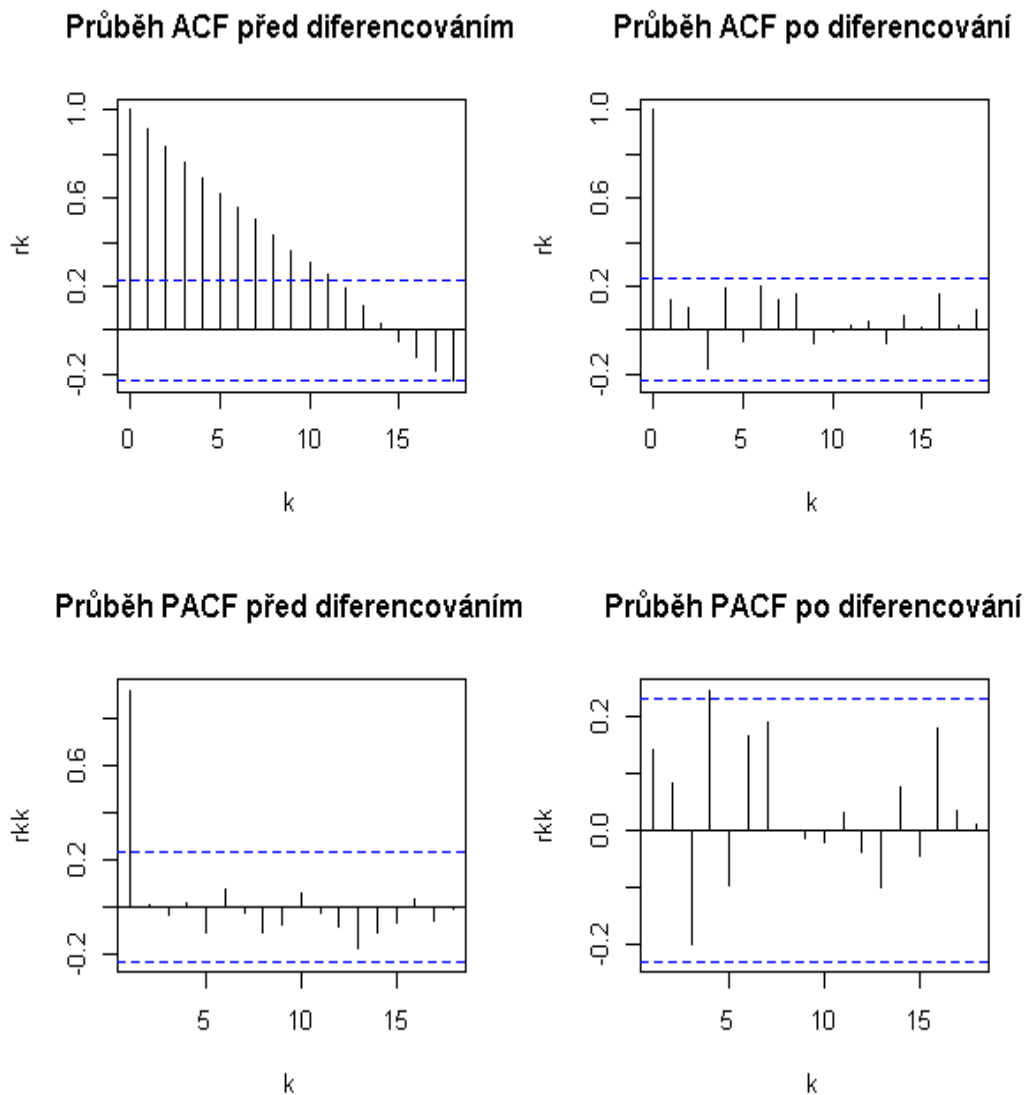
Z obr. 3.11 je zřejmé, že časová řada není stacionární, proto bude nutné řadu opět diferencovat. Programový kód pro zobrazení průběhu této řady (obr. 3.11) spolu s autokorelační a parciální autokorelační funkcí (obr. 3.12) vypadá následovně

```
> PF2 <- read.csv(„D:/PF2.csv“, header=T, dec=„.“, sep=„;“)
> plot(PF2[,2],xlim=c(0,110),ylim=c(0,1700),type=„l“)
> par(mfrow=c(2,2))
> acf(PF2[,2],main=„Průběh ACF před diferencováním“, xlab=„k“,ylab=„rk“)
> acf(diff(PF2[,2]),main=„Průběh ACF po diferencování“, xlab=„k“,ylab=„rk“)
> pacf(PF2[,2],main=„Průběh PACF před diferencováním“, xlab=„k“,ylab=„rkk“)
> pacf(diff(PF2[,2]),main=„Průběh PACF po diferencování“, xlab=„k“,ylab=„rkk“).
```

Příkaz `> par(mfrow=c(2,2))` vytvoří fiktivní tabulku, do které R umístí grafy zobrazující průběhy autokorelací a parciálních autokorelací funkcí viz obr. 3.12.

Z průběhu ACF a PACF lze vyčíst, že se podle základních předpokladů jedná o autoregresní proces 4. řádu, jelikož autokorelační funkce nemá identifikační bod a hodnota parciální autokorelační funkce překračuje dvojnásobek směrodatné odchylky odhadnutých reziduí v $k = 4$, PACF má identifikační bod $k_0 = 4$. Jedná se tedy o model AR (4), neboli autoregresní proces 4. řádu. V praxi jsou procesy takto vysokých řádů velice málo užívané, přesto však tento model odhadneme.

Obrázek 3.12 – Autokorelační a parciální autokorelační funkce



Zdroj: Vlastní zpracování dat v R

3.4.2 Odhad parametrů modelu

Odhady jednotlivých parametrů $\vartheta_1, \vartheta_2, \vartheta_3, \vartheta_4$ byly získány pomocí tohoto programového kódu

```
> PF2.fit<-arima(PF2[,2],order=c(4,0,0))
> PF2.fit
```

Poslední řádek vrátí hodnoty jednotlivých parametrů spolu se směrodatnými odchylkami. Tyto hodnoty jsou zachyceny v tabulce 3.8.

Tabulka 3.8 - Odhady parametrů

Parametr	ξ_1	ξ_2	ξ_3	ξ_4
Odhad	1.1458	-0.0435	-0.3230	0.1991
Přesnost	± 0.1178	± 0.1764	± 0.1768	± 0.1185

Zdroj: Vlastní zpracování dat v R

Je vidět, že směrodatné odchylky některých parametrů jsou relativně dost vysoké, a proto je na místě domnívat se, že zvolený model není kvalitní. Proto bude opět nutné provést diagnostiku tohoto modelu.

3.4.3 Diagnostika modelu

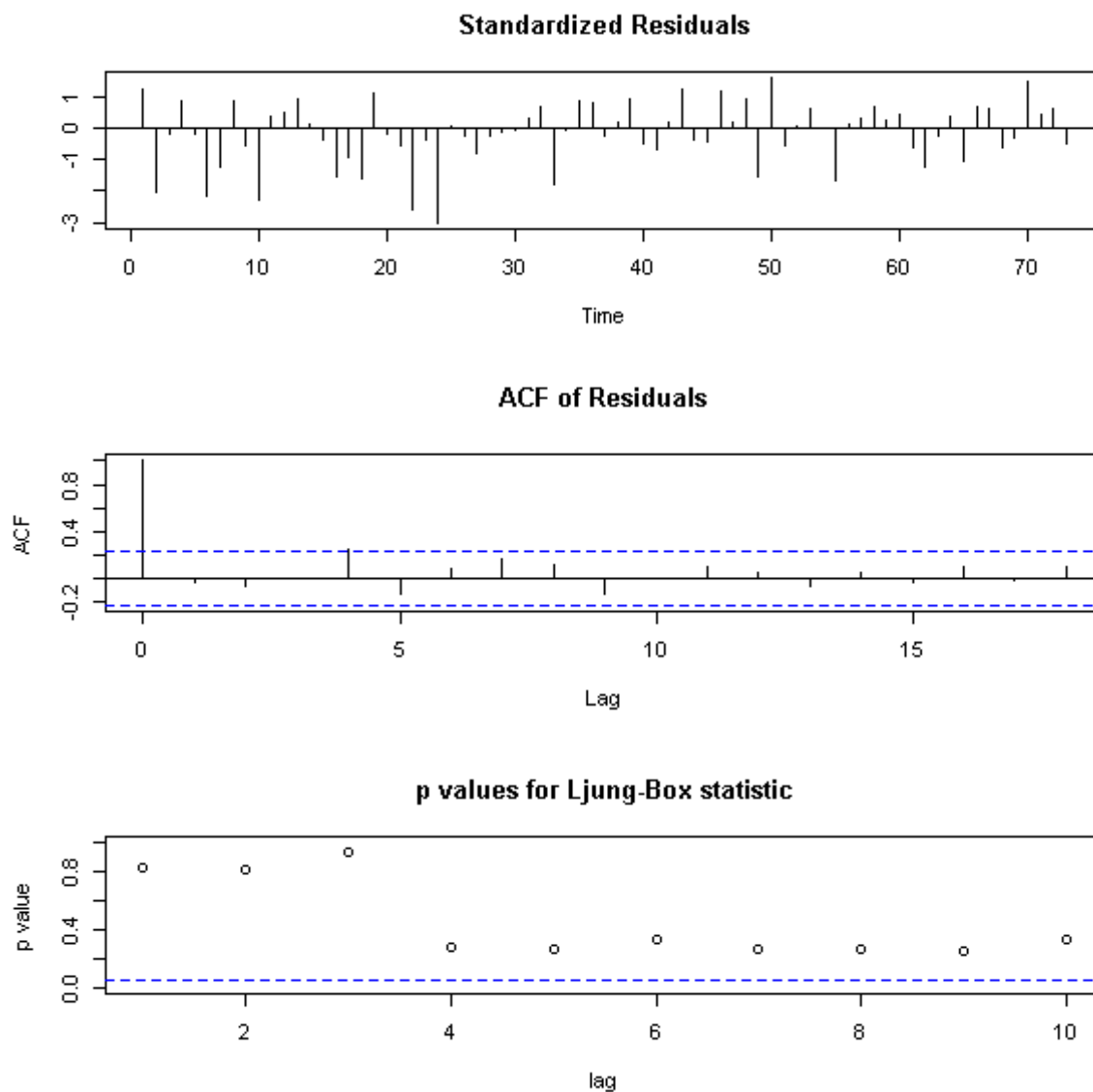
Diagnostiku provedu opět pomocí metody odhadnutých reziduí (za užití příkazu *tsdiag(PF2[,2])*). Výsledky jsou zachyceny na obr. 3.13. Z průběhu autokorelační funkce odhadnutých reziduí je vidět, že pro $k=4$ je dvojnásobek směrodatné odchylky těchto odhadů překročen (byť jen velice málo). Proto bychom mohli vyslovit domněnku o nevhodnosti zvoleného modelu. Rozhodl jsem se však ověřit kvalitu modelu ještě Portmanteau testem. Tento test vyhodnocuje souhrnné autokorelační chování odhadnutých reziduí, tedy je o málo slabším testem oproti předcházející metodě, která odhady zkoumá individuálně [6]. Testová statistika vypadá následovně

$$Q = n \sum_{k=1}^K \hat{r}_k^2(\hat{\varepsilon}) = 73 \cdot [(-0,025)^2 + (-0,067)^2 + \dots + (-0,127)^2] = 10,156,$$

kde $K = \sqrt{n} = \sqrt{73} = 8,54$, tedy 9. Testujeme nulovou hypotézu $H_0 : \rho(\hat{\varepsilon}) = 0$. Hodnotu statistiky porovnáme s kritickou hodnotou $\chi_{9-4}^2(1-\alpha)$. Z tabulky 1b v příloze B zjistíme, že kritická hodnota je 16,91896. Platí tedy $Q < 16,91896$ a proto nulovou hypotézu nelze zamítnout. Tento fakt je dále potvrzen Ljung-Boxovu statistikou, jejíž *P-values* nabývají hodnot větších, než jaká je hladina významnosti α reprezentovaná modrou přerušovanou

hranicí ve spodní části obr. 3.13, proto tedy opravdu nelze $H_0: \rho_1 = \rho_2 = \dots = \rho_m = 0$ zamítnout.

Obrázek 3.13 – Diagnostika modelu



Zdroj: Vlastní zpracování dat v R

Hodnoty autokorelační funkce reziduí, taktéž z obr. 3.13, byly získány pomocí programového kódu

```
> PF2.fit<-arima(PF2[,2],order=c(4,0,0))
> PF2.fit
> tsvdiag(PF2.fit)
```

```
> Box.test(PF2.fit$residuals,lag=9)
> PF2.fit$residuals
> acf(PF2.fit$residuals,plot=FALSE),
```

kde jejich výpočet spočívá prakticky v dosazování do následující rovnice

$$\hat{\varepsilon}_t = \hat{\theta}^{-1} \hat{\zeta}(B) y_t.$$

Model je tedy zřejmě vyhovující a má tvar

$$\hat{y}_t = (1,1458 \pm 0,11178) y_{t-1} - (0,0435 \pm 0,1764) y_{t-2} - (0,3230 \pm 0,1768) y_{t-3} + (0,1991 \pm 0,1185) y_{t-4} + \varepsilon_t$$

3.4.4 Bodová a intervalová predikce

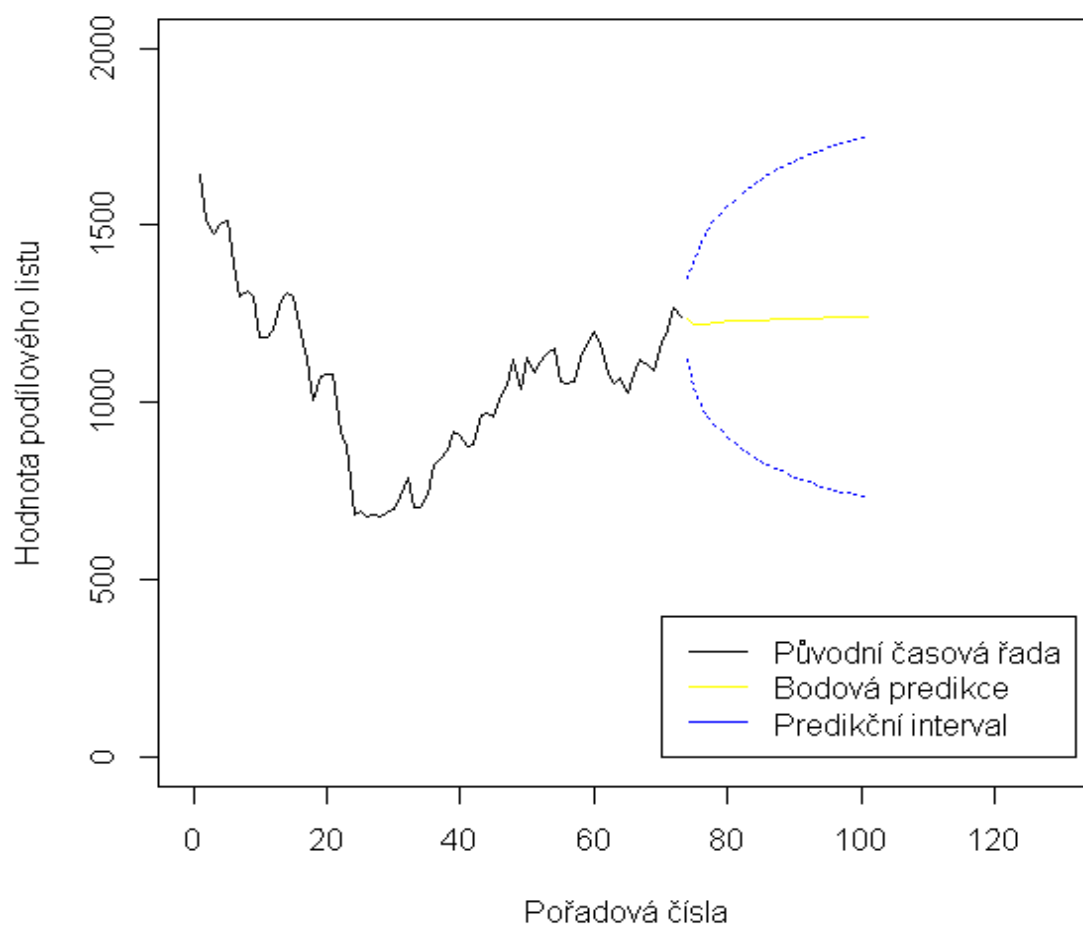
Pro predikci v této časové řadě použijí opět programového kódu

```
> PF2.fit<-arima(PF2[,2],order=c(4,0,0))
> plot(PF2[,2],type="l",xlim=c(0,100),ylim=c(0,2000),xlab="Pořadová čísla", ylab =
„Hodnota podílového listu“, main="Bodová a intervalová predikce")
> PF2.pred<-predict(PF2.fit,n.ahead=28)
> lines(PF2.pred$pred,col="yellow")
> lines(PF2.pred$pred-2*PF2.pred$se,col="blue",lty=3)
> lines(PF2.pred$pred+2*PF2.pred$se,col="blue",lty=3)
> legend(70,400,c("Původní časová řada","Bodová predikce","Predikční interval"),
lty=c(1,1,1),col=c("black","yellow","blue")).
```

Malou nevýhodou, či slabou stránkou předpovědí v tomto modelu bude fakt, že předpovídáme o 28 časových jednotek dopředu (tedy až do konce roku 2011 jako u předchozích metod). Jedná se relativně o dlouhé predikční období snižující

důvěryhodnost těchto predikcí. Období, pro která predikujeme spolu s bodovými a intervalovými predikcemi, jsou zahrnuty v tabulce 14a v příloze A. Bodové predikce a predikční intervaly jsou také znázorněny na obr. 3.14. Jak je vidět, predikční intervaly jsou dosti široké, jak jsem již předestíral na začátku kapitoly 3.6. Proto je dosti pravděpodobné, že se skutečné hodnoty sice budou pohybovat v rámci těchto predikčních intervalů, pravděpodobně však nebudou příliš korespondovat s bodovou predikcí.

Obrázek 3.14 – Bodová a intervalová předpověď



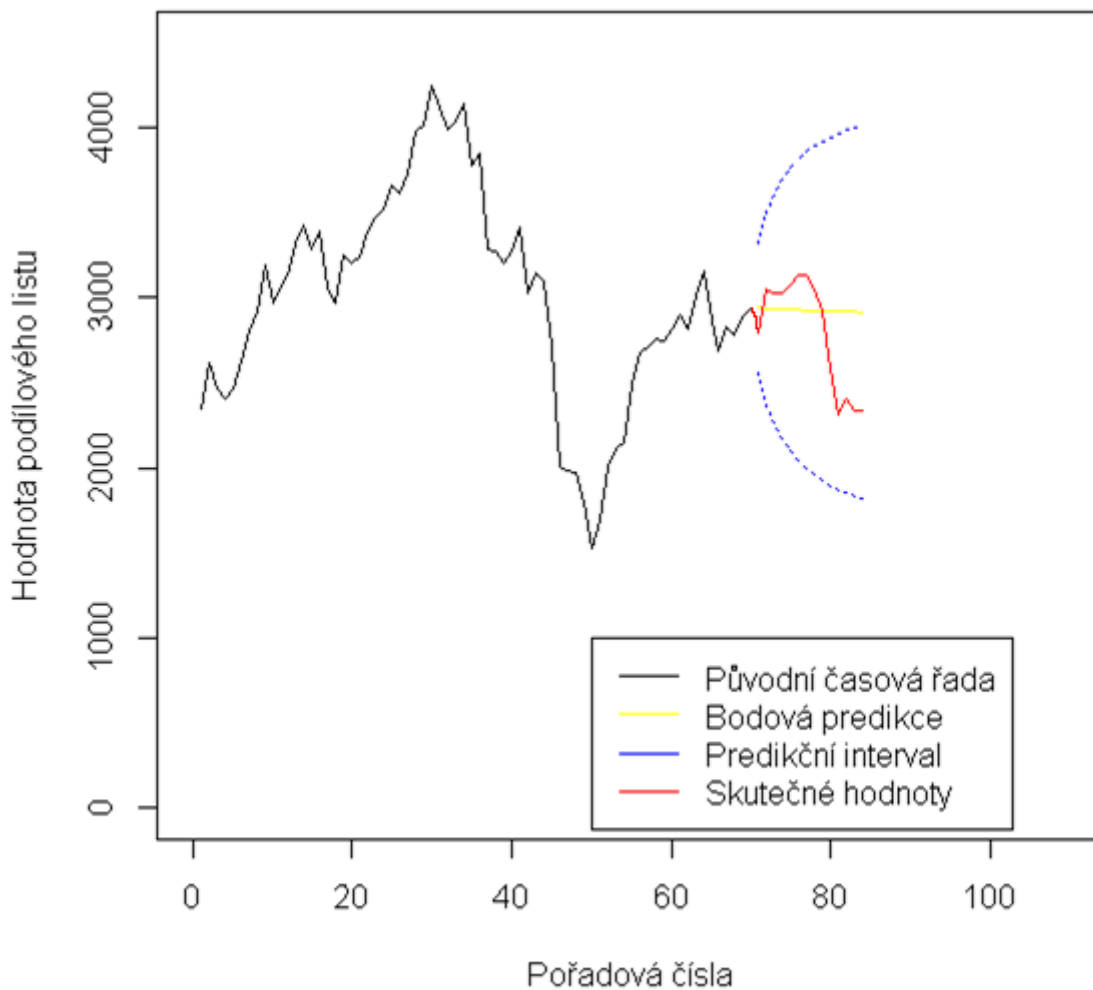
Zdroj: Vlastní zpracování dat v R

4 Porovnání se skutečně dosaženými hodnotami

4.1 Box-Jenkinsova metodologie

Obr. 4.1 indikuje určitý rozpor v prvotních tendencích a bodovou předpověď lze považovat pouze za ukazatel dalšího dlouhodobého vývoje (je vidět, že bodová predikce má klesající charakter a naznačuje tedy „medvědí trend“ - klesající tendenci). Porovnáme-li spočítané predikce se skutečnými hodnotami, dostaneme opět relativně vysoké odchylky (viz tab. 4.1). Pozitivním se však jeví to, že po dobu celého predikovaného období se skutečné hodnoty pohybují v mezích predikčních intervalů.

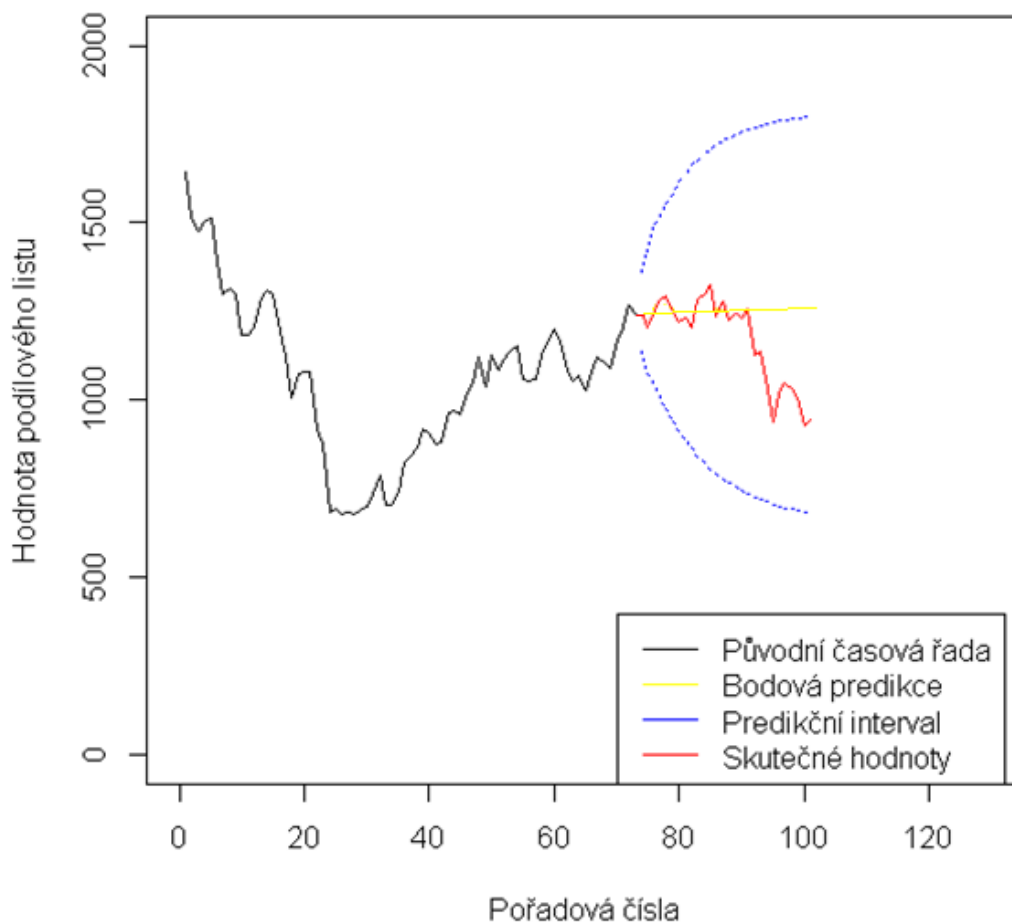
Obrázek 4.1 – Srovnání se skutečně dosaženými hodnotami PF1



Zdroj: Vlastní zpracování dat

Z obrázku 4.2 lze vypožorovat, že v prvních fázích vývoje časové řady byla bodová predikce dosti přesná a působila téměř jako odhad tendencí tohoto fondu (viz žlutá křivka na obr. 4.2). Je však také vidět, že kvalita předpovědí se v čase vytrácí a bodová předpověď na konci sledovaného období je dosti nepřesná. Konkrétní nepřesnost předpovědi je zachycena v tabulce 4.1.

Obrázek 4.2 – Srovnání se skutečně dosaženými hodnotami PF2



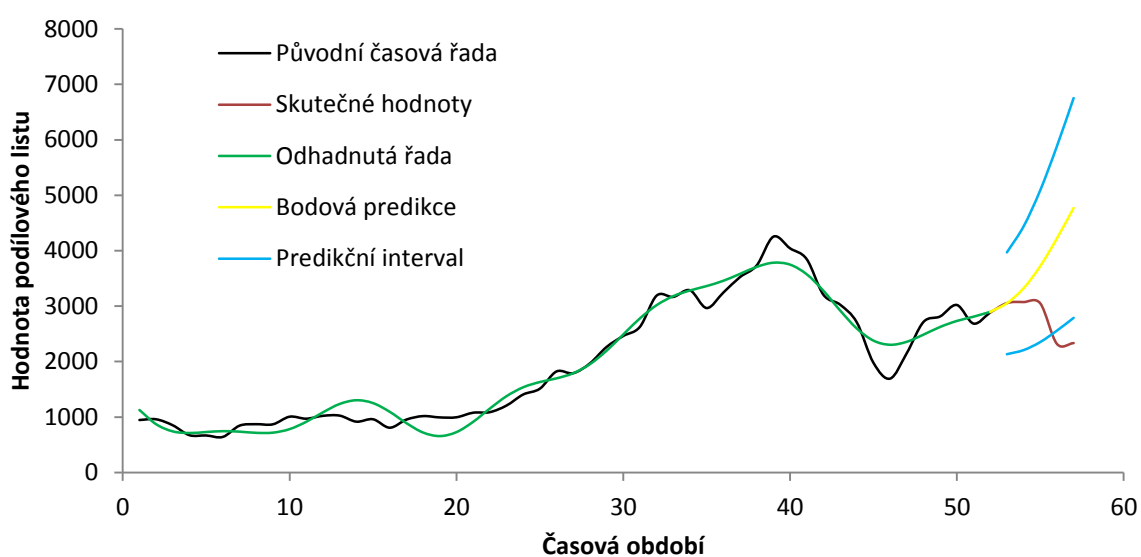
Zdroj: Vlastní zpracování dat

4.2 Dekompoziční přístup

Spektrální analýza u časové řady zachycující vývoj PF1 odhalila hned 4 významné periody, o jejichž kvalitě vypovídá samotná odhadnutá řada na obr. 4.3. Relativně malé šířky predikčních intervalů napovídaly tomu, že by tento model mohl relativně dobře vystihnout budoucí chování, což se nakonec dostatečně neukázalo. Z obr. 4.3 je vidět, že se skutečné hodnoty pohybují mimo predikční intervaly (na konci predikovaného období). Skutečná hodnota 2334,17 Kč se liší od odhadované 4770,61 Kč poměrně značně, je tedy zřejmé, že i navzdory 4 nalezených významným period, se nepodařilo predikovat vývoj hodnoty podílového listu dostatečně dobře. Informace z modelu předestírala prudký nárůst hodnoty pro rok 2011, což by de facto znamenalo signál k udržení pozice pro investora (tedy ani nenakoupit ani neprodat podílové listy). Je zřejmé, že tento signál nekoresponduje se skutečným chováním hodnoty podílového listu, proto by pro investora byl signálem klamným.

Jak je vidět, v predikované části postupuje bodová predikce rychle směrem k vyšším hodnotám podílového listu. Lze předpokládat, že do tohoto průběhu významně zasáhla perioda, která se vytvořila v původních datech na základě ekonomické recese (zachyceno hodnotami s pořadovými čísly 30-52). Je tedy pravděpodobné, že byl celý model negativně ovlivněn touto událostí, a proto došlo k tak značným odchylkám predikovaných a skutečných hodnot.

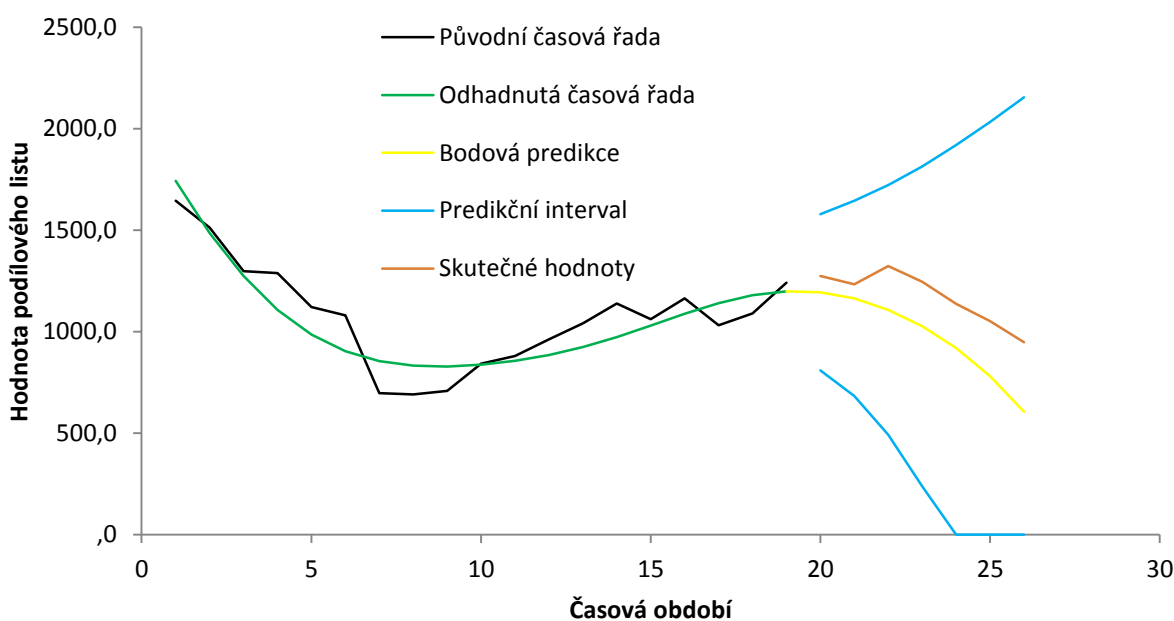
Obrázek 4.3 - Srovnání se skutečně dosaženými hodnotami PF1



Zdroj: Vlastní zpracování dat

Z obr. 4.4 můžeme vypořádat, že bodová predikce jakoby kopírovala tendence skutečného vývoje. Po celé predikované období nedošlo k překročení predikčního intervalu, což je do značné míry pozitivní. Ukazuje se tedy, že klesající trend této řady byl dobře vystižen. Přestože se nepodařilo vytvořit predikce přesněji, tedy rozdíl predikované a skutečné hodnoty na konci predikovaného období je poměrně značný, hodnotím tento model přínosně z pohledu předpovídané tendence vývoje sledované hodnoty. Méně pozitivním faktem však zůstává to, že predikovaný model opět dával určitý signál k negativním výsledkům. Skutečné hodnoty by však tento signál potvrdily jen částečně.

Obrázek 4.4 – Srovnání se skutečně dosaženými hodnotami PF2



Zdroj: Vlastní zpracování dat

4.3 Závěrečné shrnutí výsledků

Následující tabulka 4.1 shrnuje výsledky obou přístupů (Box-Jenkins, dekompoziční přístup) k predikování budoucího vývoje hodnot vybraných podílových listů.

Tabulka 4.1 – Zpracování odchylek jednotlivých přístupů

	Podílový fond č. 1			Podílový fond č. 2		
	Predikce k 30.12.11	Skutečná hodnota	Odchylka od skutečné hodnoty %	Predikce k 30. 12. 11	Skutečná hodnota	Relativní odchylka %
Dekompoziční přístup	4770,61	2 334,17	104,38	605,29	946,59	36,05
Box- Jenkinsova metodologie	2917,25	2 334,17	24,98	1242,45	946,59	31,25

Zdroj: Vlastní zpracování dat

Relativní odchylky se u jednotlivých přístupů liší. Nejlepší předpovědi, resp. ty s nejmenší odchylkou, poskytla Box-Jenkinsova metodologie, kde např. pro 1. podílový fond činí odchylka necelých 25% a pro 2. podílový fond činí odchylka zhruba 30%. Tyto odchylky predikovaných hodnot od hodnot skutečných nejsou pozitivní. Toto může být způsobeno převážně citlivostí hodnoty podílového listu na prakticky jakékoliv vnější podmínky. Jednoduše řečeno, ani studie všech složek dekompozice nemusí dát relevantní výsledky, jelikož i špatný výrok politika může ze dne na den vytvořit obrovský propad cen, bez podstatnější příčiny, která by se právě periodickou složkou dala vysvětlit. Následné „zotavení“ vývoje sledované hodnoty může trvat i rok.

V tabulce 4.1 jsou tučnými rámečky zviditelněny ty odchylky predikcí od skutečných hodnot, které byly nejvyšší.

Jistým zklamáním pro mne je opravdu vysoká odchylka u spektrální analýzy pro 1. podílový fond, která činí téměř 60%. Při grafickém znázornění šířky predikčních intervalů jsem vkládal velkou naději v tento model. Ukázal se však pravý opak.

Závěr

Cílem této práce bylo použít přístupy k analýze časových řad k predikování budoucího vývoje hodnot dvou vybraných podílových listů. Predikování ekonomických, ale i jiných, veličin je obecně složitou procedurou s výsledky zatíženými velkou nejistotou. Do vývoje takovýchto veličin často zasahuje široké spektrum faktorů od politického vývoje až po špatné (a často falešné) výroky vysoce postavených manažerů či politiků. V současné době nelze tyto vlivy zahrnout do matematického modelování, jelikož jsou to faktory zcela nepředvídatelné a označované za tzv. „The Black Swan“, jak popisuje Nassim Nicholas ve stejnojmenné knize. Podílové fondy, resp. vývoj hodnoty podílových listů jsou právě těmito nepředvídatelnými faktory velice často zasaženy. Přestože se nám může podařit vytvořit korektní matematický model, který by vystihoval alespoň přibližně (což v praxi často stačí) charakter dané časové řady, jediný výše zmíněný faktor lehce tento model naruší, či zcela znehodnotí výsledky jeho zpracování. Určitým handicapem, však zcela logickým, je určité omezení na počet období, pro která se snažíme predikovat vývoj dané veličiny. Je zřejmé, že čím více takovýchto období volíme, tím s menší pravděpodobností tyto predikce nastanou.

Rozsahy obou časových řad byly úmyslně zkráceny, jelikož celkový součet dat činil cca 3500 pro PF1 a téměř 1000 pro PF2. Selektce dat pro analýzu byla provedena s ohledem na volbu predikčního období, tedy rok 2011 a na co nejmenší ztrátu informace o dané časové řadě.

Jako první jsem ve své práci zvolil dekompoziční přístup, jehož účelem je rozložit sledovanou časovou řadu do několika separovaných složek. První analyzovanou složkou byla složka trendová, jež se hodnotila z několika hledisek. Pro každý z analyzovaných fondů byla nalezena vhodná matematická funkce vystihující její dlouhodobé chování. Druhou analyzovanou složkou byla složka periodická, k jejíž analýze byl použit model skrytých period. Při zanesení obou těchto složek do celkového modelu byly získány hodnoty bodových a intervalových predikcí. Při závěrečném porovnávání těchto hodnot s hodnotami skutečně dosaženými však byly výsledky poměrně nekvalitní a nemohli by tak sloužit ke kvalitnímu zkoumání těchto časových řad.

Jako druhý přístup jsem zvolil Box-Jenkinsovu metodologii, která byla vyvinuta právě s ohledem na to, že v hospodářské praxi je většina časových řad popisujících ekonomické veličiny velice volatilních (kolísavých) a v čase tedy mění svůj charakter neboli chování. Na základě průběhů autokorelačních a parciálních autokorelačních funkcí byly identifikovány

jednotlivé modely a za užití matematického software R byly spočteny bodové predikce a predikční intervaly. Jak je vidět z obrázku 4.1, první podílový fond byl relativně dobře odhadnut s ohledem na tendence svého vývoje v prvních dvou třetinách predikovaného období. Jelikož získaná data byla selektována s jinými časovými rozestupy (oproti dekompozičnímu přístupu), třetí třetina predikovaného období již neodpovídala predikovaným hodnotám (pořád se však hodnoty pohybovaly v predikčních intervalech). Přiřazuji tuto skutečnost převážně vzdálenosti těchto období od referenčního bodu- tj. datumu posledního známého údaje o hodnotě podílového listu. Na obr. 4.2, který zachycuje srovnání predikovaných a skutečných hodnot pro PF2 je vidět, že se predikované hodnoty nechovaly v souladu se skutečnými hodnotami. Bodová predikce dokázala jen částečně vystihnout počáteční chování dané časové řady, však predikce ke 30.12.2011 dosáhla odchylky více než 30%. Takovéto odchylky mohou způsobit milionové ztráty z investic, a proto nepovažuji tyto predikce za příliš kvalitní.

Je zřejmé, že analýza časových řad takovéhoho charakteru je silně závislá na politické, ekonomické a sociální stabilitě, která prakticky vytváří hodnoty cenných papírů. Jako hlavní příčinu neúspěchu obou zvolených přístupů vidím ekonomickou recesy od roku 2008, která nestabilitu zmíněných segmentů dále vystupňovala.

LITERATURA

- [1] LIŠKA,V.,GAZDA,J., *Kapitálové trhy a kolektivní investování*, Professional Publishing, 2004,
- [2] STEIGAUFG, S. *Fondy: Jak vydělávat pomocí fondů*. 1.vyd. Praha: Grada Publishing, 2003. ISBN 80-247-0247-9.
- [3] BENJAMIN,G., *Inteligentní investor*, Grada, ISBN: 978-80-247-1792-0
- [4] SCHILLER,R., *Investiční horečka*,Grada, 2010, ISBN: 978-80-247-2482-9
- [5] KOMISE PRO CENNÉ PAPÍRY, *Umění investovat*, str. 10
- [6] CIPRA, T. *Analýza časových řas s aplikacemi v ekonomii*. Praha: SNTL, 1986.
- [7] KŘIVÝ, IVAN. *Analýza časových řad*. distanční opora, Ostrava: Ostravská univerzita.
- [8] BOX, G. E.P., JENKINS, G. M. *Time series analysis forecasting and control*. Oakland, 1976. 76-8718.
- [9] VOŠVRDA, M. S. *Časové řady s programem STATISTICA*. Praha : StatSoft CR s.r.o., 2003. ISBN 978-80-904033-2-1.
- [10] Time Series Analysis with R [online], dostupné z: http://www.statoek.wiso.uni-goettingen.de/veranstaltungen/zeitreihen/sommer03/ts_r_intro.pdf
- [11] CRYER, D., J., CHAN, K., S. *Time Series Analysis with Applications in R*, Springer, 2ndedition, p. 175
- [12] RAWLINGS, O. J., PANTULA, S.G., DICKEY, D. A. *Applied Regression Analysis: A Research Tool*, 2nd Edition, ISBN 0-387-98454-2, p. 222
- [13] TSAY, R., S., *Analysis of financial time series*, Wiley, 3rd edition, ISBN 978-0-470-41435-4, p. 32
- [14] ARTL,J., ARTLOVÁ, M., RUBLÍKOVÁ, E., *Analýza ekonomických časových řad s příklady*, Vysoká škola ekonomická v Praze, 2002, učební materiál
- [15] VEČEŘA, J., *Role investičních společností a fondů na českém kapitálovém trhu*, Bakalářská práce, Moravská vysoká škola Olomouc, 2010
- [16] MARQUES de SÁ,J.,P., *Applied Statistics using SPSS, STATISTICA and MATLAB*, Springer, 2003, ISBN 3-540-01156-0

Příloha A

Tabulka 1a - Pořadové hodnoty + odhadnuté hodnoty polynomu 2. stupně PF1

Datum	Pořadové číslo	*Odhadnuté hodnoty		Datum	Pořadové číslo	*Odhadnuté hodnoty
31.12.97	1	231,60	<i>t</i>	29.12.06	37	2633,52
31.03.98	2	317,22		30.03.07	38	2680,26
30.06.98	3	401,76		29.06.07	39	2725,92
30.09.98	4	485,22		28.09.07	40	2770,50
31.12.98	5	567,60		31.12.07	41	2814,00
31.03.99	6	648,90		31.03.08	42	2856,42
30.06.99	7	729,12		30.06.08	43	2897,76
30.09.99	8	808,26		30.09.08	44	2938,02
30.12.99	9	886,32		31.12.08	45	2977,20
31.03.00	10	963,30		31.03.09	46	3015,30
30.06.00	11	1039,20		30.06.09	47	3052,32
29.09.00	12	1114,02		30.09.09	48	3088,26
19.12.00	13	1187,76		31.12.09	49	3123,12
30.03.01	14	1260,42		31.03.10	50	3156,90
<i>t</i> 29.06.01	15	1332,00		30.06.10	51	3189,60
27.09.01	16	1402,50		30.09.10	52	3221,22
31.03.03	22	1802,82				
30.06.03	23	1865,76				
30.09.03	24	1927,62				
31.12.03	25	1988,40				
31.03.04	26	2048,10				
30.06.04	27	2106,72				
30.09.04	28	2164,26				
31.12.04	29	2220,72				
31.03.05	30	2276,10				
30.06.05	31	2330,40				
30.09.05	32	2383,62				
30.12.05	33	2435,76				
31.03.06	34	2486,82				
30.06.06	35	2536,80				
29.09.06	36	2585,70				

Zdroj: Vlastní zpracování dat

$$*) \hat{y}_t = 144,9 + 87,24t - 0,547t^2$$

Pozn: *t*..... časové okamžiky, pro něž jsou dostupná historická data

Tabulka 2a - Pořadové hodnoty + odhadnuté hodnoty polynomu 3. stupně PF2

	Datum	Pořadové číslo	* Odhadnuté hodnoty
t	31.10.2007	1	1740,40
	31.12.2007	2	1495,88
	29.02.2008	3	1296,66
	30.04.2008	4	1138,96
	30.06.2008	5	1019,00
	29.08.2008	6	933,00
	31.10.2008	7	877,18
	31.12.2008	8	847,76
	27.02.2009	9	840,96
	30.04.2009	10	853,00
	30.06.2009	11	880,10
	31.08.2009	12	918,48
	30.10.2009	13	964,36
	31.12.2009	14	1013,96
	26.02.2010	15	1063,50
	30.04.2010	16	1109,20
	30.06.2010	17	1147,28
	31.08.2010	18	1173,96
	28.10.2010	19	1185,46

Zdroj: Vlastní zpracování dat

*) $\hat{y}_t = 2034,75 - 319,48t + 26,43t^2 - 0,63t^3, \quad t = 1 \dots n$

Tabulka 3a - Očištění od trendové složky PF1

Časový okamžik t	Odpovídající datum	Původní časová řada y_t	*Odhadnutá časová řada -polynom 2. stupně \hat{y}_t	$y_t^{(k)} = y_t - \hat{y}_t$
1	31.12.1997	947,13	231,60	715,53
2	31.03.1998	959,75	317,22	642,53
3	30.06.1998	851,35	401,76	449,59
4	30.09.1998	673,04	485,22	187,82
5	31.12.1998	668,15	567,60	100,55
6	31.03.1999	642,37	648,90	-6,53
7	30.06.1999	845,29	729,12	116,17
8	30.09.1999	871,50	808,26	63,24
9	30.12.1999	869,42	886,32	-16,90
10	31.03.2000	1006,61	963,30	43,31
11	30.06.2000	970,86	1039,20	-68,34
12	29.09.2000	1022,50	1114,02	-91,52
13	19.12.2000	1026,00	1187,76	-161,76
14	30.03.2001	917,02	1260,42	-343,40
15	29.06.2001	960,71	1332,00	-371,29
16	27.09.2001	808,17	1402,50	-594,33
17	31.12.2001	954,29	1471,92	-517,63
18	29.03.2002	1017,52	1540,26	-522,74
19	28.06.2002	993,67	1607,52	-613,85
20	30.09.2002	996,34	1673,70	-677,36
21	31.12.2002	1078,22	1738,80	-660,58
22	31.03.2003	1088,61	1802,82	-714,21
23	30.06.2003	1207,17	1865,76	-658,59
24	30.09.2003	1406,62	1927,62	-521,00
25	31.12.2003	1511,52	1988,40	-476,88
26	31.03.2004	1822,52	2048,10	-225,58
27	30.06.2004	1791,32	2106,72	-315,40
28	30.09.2004	1967,42	2164,26	-196,84
29	31.12.2004	2263,76	2220,72	43,04
30	31.03.2005	2462,29	2276,10	186,19
31	30.06.2005	2626,91	2330,40	296,51
32	30.09.2005	3189,17	2383,62	805,55
33	30.12.2005	3166,64	2435,76	730,88
34	31.03.2006	3286,44	2486,82	799,62
35	30.06.2006	2966,08	2536,80	429,28
36	29.09.2006	3244,75	2585,70	659,05
37	29.12.2006	3524,10	2633,52	890,58
38	30.03.2007	3734,06	2680,26	1053,80
39	29.06.2007	4253,10	2725,92	1527,18
40	28.09.2007	4042,69	2770,50	1272,19
41	31.12.2007	3846,98	2814,00	1032,98
42	31.03.2008	3206,25	2856,42	349,83

43	30.06.2008	3026,69	2897,76	128,93
44	30.09.2008	2708,83	2938,02	-229,19
45	31.12.2008	1978,44	2977,20	-998,76
46	31.03.2009	1693,52	3015,30	-1321,78
47	30.06.2009	2154,28	3052,32	-898,04
48	30.09.2009	2714,83	3088,26	-373,43
49	31.12.2009	2815,11	3123,12	-308,01
50	31.03.2010	3021,12	3156,90	-135,78
51	30.06.2010	2685,60	3189,60	-504,00
52	30.09.2010	2885,76	3221,22	-335,46

Zdroj: Vlastní zpracování dat z www.ing.cz

Tabulka 4a - Hodnoty pro periodogram PF1

Pořadí j	Frekvence ω_j	** \hat{a}_j	*** \hat{b}_j	**** \hat{v}_j
1	0,12	-146,10	264,47	45644,80
2	0,24	230,47	50,07	27811,71
3	0,36	-31,57	-9,98	548,13
4	0,48	80,25	56,16	4797,00
5	0,60	27,00	16,66	503,28
6	0,72	86,47	18,03	3901,07
7	0,85	32,06	-21,63	747,85
8	0,97	49,10	-0,75	1205,69
9	1,09	35,38	-18,58	798,48
10	1,21	36,62	-28,71	1082,64
11	1,33	7,12	-20,95	244,80
12	1,45	23,78	-17,79	440,99
13	1,57	4,99	-19,62	204,92
14	1,69	10,87	-13,98	156,80
15	1,81	8,59	-6,12	55,62
16	1,93	19,20	-21,41	413,51
17	2,05	-0,80	-12,20	74,74
18	2,17	18,47	-17,87	330,24
19	2,29	-6,96	-23,74	306,01
20	2,42	-3,02	-6,34	24,66
21	2,54	2,78	-1,41	4,86
22	2,66	17,06	-10,51	200,75
23	2,78	3,27	-21,15	229,01
24	2,90	1,04	-11,33	64,73
25	3,02	7,53	-14,87	138,91
26	3,14	0,18	-22,91	45644,80

Zdroj: Vlastní zpracování dat z www.ing.cz

$$*) \omega_j = \frac{2\pi j}{n}, \quad **) \hat{a}_j = \frac{2}{n} \sum_{t=1}^n y_t \sin \omega_j t, \quad ***) \hat{b}_j = \frac{2}{n} \sum_{t=1}^n y_t \cos \omega_j t, \quad ****) \hat{v}_j = \frac{1}{2} (\hat{a}_j^2 + \hat{b}_j^2)$$

Tabulka 5a – Nově odhadnutý model obsahující trendovou i periodickou složku (PF1)

t	Původní řada y_t	$y_t^{(o)} = y_t - \hat{y}_t$	Nový model *
1	947,13	715,53	1127,36
2	959,75	642,53	870,50
3	851,35	449,59	741,67
4	673,04	187,82	712,44
5	668,15	100,55	730,37
6	642,37	-6,53	745,97
7	845,29	116,17	736,72
8	871,50	63,24	715,48
9	869,42	-16,90	718,93
10	1006,61	43,31	782,46
11	970,86	-68,34	914,68
12	1022,50	-91,52	1085,71
13	1026,00	-161,76	1236,18
14	917,02	-343,40	1303,41
15	960,71	-371,29	1252,29
16	808,17	-594,33	1095,27
17	954,29	-517,63	890,99
18	1017,52	-522,74	721,29
19	993,67	-613,85	656,95
20	996,34	-677,36	728,14
21	1078,22	-660,58	913,45
22	1088,61	-714,21	1152,53
23	1207,17	-658,59	1375,78
24	1406,62	-521,00	1536,79
25	1511,52	-476,88	1631,85
26	1822,52	-225,58	1697,84
27	1791,32	-315,40	1790,24
28	1967,42	-196,84	1953,06
29	2263,76	43,04	2195,86
30	2462,29	186,19	2489,08
31	2626,91	296,51	2779,59
32	3189,17	805,55	3017,59
33	3166,64	730,88	3180,75
34	3286,44	799,62	3282,82
35	2966,08	429,28	3362,43
36	3244,75	659,05	3457,89
37	3524,10	890,58	3581,69
38	3734,06	1053,80	3708,41
39	4253,10	1527,18	3783,43
40	4042,69	1272,19	3748,92
41	3846,98	1032,98	3574,59
42	3206,25	349,83	3277,91
43	3026,69	128,93	2922,93
44	2708,83	-229,19	2597,47
45	1978,44	-998,76	2378,80
46	1693,52	-1321,78	2303,85
47	2154,28	-898,04	2357,82
48	2714,83	-373,43	2486,48
49	2815,11	-308,01	2625,60
50	3021,12	-135,78	2733,38
51	2685,60	-504,00	2810,20
52	2885,76	-335,46	2896,64

Zdroj: Vlastní zpracování dat

$$*) \hat{y}_t = 3539,65 - 238,74t + 5,11t^2 - 923,86 \sin(0,12t) - 1476,69 \cos(0,12t) + 233,37 \sin(0,24t) - 641,48 \cos(0,24t) - 24,02 \sin(0,48t) + 150,35 \cos(0,48t) - 155,87 \sin(0,72t) - 70,85 \cos(0,72t)$$

Tabulka 6a – Hodnoty bodových predikcí a predikčních intervalů pro PF1

		Datum	Bodová predikce \hat{y}_T^p	Predikční interval
T	53	30.12.2010	3051,71	< 2133,13 ; 3970,28 >
	54	29.3.2011	3322,69	< 2207,41 ; 4437,98 >
	55	28.6.2011	3721,95	< 2352,75 ; 5091,15 >
	56	30.9.2011	4221,97	< 2558,91 ; 5885,03 >
	57	30.12.2011	4770,61	< 2788,76 ; 6752,46 >

Zdroj: Vlastní zpracování dat

Tabulka 7a - Očištění od trendové složky PF2

Časový okamžik t	Odpovídající datum	Původní řada y_t	Odhadnutá časová řada -polynom 3. stupně \hat{y}_t	$y_t^{(k)} = y_t - \hat{y}_t$
1	31.10.2007	1 644,53	1740,40	-95,87
2	31.12.2007	1 511,81	1495,88	15,93
3	29.02.2008	1 298,11	1296,66	1,45
4	30.04.2008	1 287,40	1138,96	148,44
5	30.06.2008	1 120,62	1019,00	101,62
6	29.08.2008	1 080,45	933,00	147,45
7	31.10.2008	695,91	877,18	-181,27
8	31.12.2008	689,64	847,76	-158,12
9	27.02.2009	707,02	840,96	-133,94
10	30.04.2009	841,92	853,00	-11,08
11	30.06.2009	878,86	880,10	-1,24
12	31.08.2009	961,7	918,48	43,22
13	30.10.2009	1 039,65	964,36	75,29
14	31.12.2009	1 138,10	1013,96	124,14
15	26.02.2010	1 060,38	1063,50	-3,12
16	30.04.2010	1 162,67	1109,20	53,47
17	30.06.2010	1 030,63	1147,28	-116,65
18	31.08.2010	1 088,50	1173,96	-85,46
19	28.10.2010	1 240,10	1185,46	54,64

Zdroj: Vlastní zpracování dat

Tabulka 8a - Hodnoty pro periodogram PF2

Pořadí j	Frekvence ω_j	\hat{a}_j	\hat{b}_j	\hat{v}_j
1	0,33	-3,69	14,49	111,77
2	0,66	52,05	-89,22	5 334,54
3	0,99	-38,21	14,90	841,21
4	1,32	21,93	44,65	1 237,29
5	1,65	1,37	-7,26	27,26
6	1,98	-6,88	27,12	391,30
7	2,31	-1,85	30,80	476,09
8	2,64	-27,04	-0,74	365,83
9	2,97	-43,75	20,68	1 170,80

Zdroj: Vlastní zpracování dat

$$*) \omega_j = \frac{2\pi j}{n}, \quad **) \hat{a}_j = \frac{2}{n} \sum_{t=1}^n y_t \sin \omega_j t, \quad ***) \hat{b}_j = \frac{2}{n} \sum_{t=1}^n y_t \cos \omega_j t, \quad ****) \hat{v}_j = \frac{1}{2} (\hat{a}_j^2 + \hat{b}_j^2)$$

Tabulka 9a - Výpočty pro model skrytých period a celkový model (PF2)

t	Původní řada	$y_t^{(o)} = y_t - \hat{y}_t$	Nový model *
1	1 644,53	-95,87	1 742,63
2	1 511,81	15,93	1 485,28
3	1 298,11	1,45	1 273,31
4	1 287,40	148,44	1 106,93
5	1 120,62	101,62	984,75
6	1 080,45	147,45	902,74
7	695,91	-181,27	854,40
8	689,64	-158,12	831,93
9	707,02	-133,94	827,89
10	841,92	-11,08	836,84
11	878,86	-1,24	856,01
12	961,7	43,22	884,91
13	1 039,65	75,29	923,98
14	1 138,10	124,14	972,85
15	1 060,38	-3,12	1 028,91
16	1 162,67	53,47	1 086,87
17	1 030,63	-116,65	1 139,40
18	1 088,50	-85,46	1 178,66
19	1 240,10	54,64	1 198,01

Zdroj: Vlastní zpracování dat

*) $\hat{y}_t = 2030,28 - 322,94t + 26,43t^2 - 0,62t^3 - 3,68 \sin(0,66t) + 14,85 \cos(0,66t)$.

Tabulka 10a - Hodnoty bodových predikcí a predikčních intervalů pro PF2

		Datum	Bodová predikce \hat{y}_T^p	Predikční intervaly
T	20	29.12.2010	1 193,27	< 808,42 ; 1578,11 >
	21	28.2.2011	1 162,87	< 682,03 ; 1643,72 >
	22	29.4.2011	1 107,00	< 492,01 ; 1721,99 >
	23	30.6.2011	1 025,91	< 238,05 ; 1813,77 >
	24	31.8.2011	918,28	< 0 ; 1918,40 >
	25	31.10.2011	780,22	< 0 ; 2033,16 >
	26	30.12.2011	605,29	< 0 ; 2153,28 >

Zdroj: Vlastní zpracování dat

T..... predikovaná období

Tabulka 11a- Selektovaný soubor pro Box-Jenkinsovu metodologii k PF1

Datum	Hodnoty časové řady	Datum	Hodnoty časové řady	Datum	Hodnoty časové řady	Datum	Hodnoty časové řady	Datum	Hodnoty časové řady
31.01.05	2347,96	31.03.06	3286,44	31.05.07	4009,00	31.07.08	3147,73	30.09.09	2714,83
28.02.05	2617,20	28.04.06	3387,29	29.06.07	4253,10	29.08.08	3094,43	30.10.09	2756,32
31.03.05	2462,29	31.05.06	3061,93	31.07.07	4116,81	30.09.08	2708,83	30.11.09	2752,94
29.04.05	2407,89	30.06.06	2966,08	31.08.07	3993,20	31.10.08	2013,93	31.12.09	2815,11
31.05.05	2462,64	31.07.06	3252,17	28.09.07	4042,69	28.11.08	1988,61	29.01.10	2894,83
30.06.05	2626,91	31.08.06	3206,91	31.10.07	4145,92	31.12.08	1978,44	26.02.10	2817,29
29.07.05	2809,72	29.09.06	3244,75	30.11.07	3783,66	30.01.09	1797,21	31.03.10	3021,12
31.08.05	2920,36	31.10.06	3396,70	31.12.07	3846,98	27.02.09	1530,39	30.04.10	3154,08
30.09.05	3189,17	30.11.06	3470,24	31.01.08	3291,44	31.03.09	1693,52	31.05.10	2867,64
31.10.05	2975,84	29.12.06	3524,10	29.02.08	3277,65	30.04.09	2002,65	30.06.10	2685,60
30.11.05	3066,17	31.01.07	3658,02	31.03.08	3206,25	29.05.09	2113,31	30.07.10	2827,99
30.12.05	3166,64	28.02.07	3616,10	30.04.08	3275,20	30.06.09	2154,28	31.08.10	2787,28
31.01.06	3326,60	30.03.07	3734,06	30.05.08	3413,39	31.07.09	2494,12	30.09.10	2885,76
28.02.06	3425,85	30.04.07	3977,91	30.06.08	3026,69	31.08.09	2680,21	28.10.10	2 939,18

Zdroj: Vlastní zpracování dat z www.ing.cz

Tabulka 12a – Predikované hodnoty pro PF1 (Box-Jenkins)

	Pořadové číslo	Odpovídající datum	Bodová predikce	Predikční interval
T	71	30.11.2010	2941,542	< 2555,442 ; 3327,642 >
	72	31.12.2010	2938,783	< 2377,545 ; 3500,020 >
	73	31.01.2011	2936,212	< 2258,724 ; 3613,700 >
	74	28.02.2011	2933,817	< 2169,635 ; 3698,000 >
	75	31.03.2011	2931,586	< 2099,449 ; 3763,724 >
	76	29.04.2011	2929,508	< 2042,601 ; 3816,415 >
	77	31.05.2011	2927,571	< 1995,734 ; 3859,409 >
	78	30.06.2011	2925,767	< 1956,621 ; 3894,914 >
	79	29.07.2011	2924,087	< 1923,685 ; 3924,488 >
	80	31.08.2011	2922,521	< 1895,763 ; 3949,279 >
	81	30.09.2011	2921,062	< 1871,966 ; 3970,159 >
	82	31.10.2011	2919,793	< 1851,597 ; 3987,810 >
	83	30.11.2011	2918,437	< 1834,102 ; 4002,773 >
	84	30.12.2011	2917,258	< 1819,031 ; 4015,485 >

Zdroj: Vlastní zpracování dat v R

Tabulka 13a - Selektovaná data pro Box-Jenkinsovu metodologii k PF2

Datum	Hodnoty časové řady	Datum	Hodnoty časové řady	Datum	Hodnoty časové řady	Datum	Hodnoty časové řady
31.10.07	1644,53	14.08.08	1078,27	29.05.09	920,66	15.03.10	1126,36
15.11.07	1513,35	29.08.08	1080,45	16.06.09	909,76	31.03.10	1156,60
30.11.07	1474,27	15.09.08	921,25	30.06.09	878,86	15.04.10	1197,86
14.12.07	1500,37	30.09.08	879,54	15.07.09	883,03	30.04.10	1162,67
31.12.07	1511,81	15.10.08	685,48	31.07.09	963,68	14.05.10	1086,34
15.01.08	1393,66	31.10.08	695,91	14.08.09	968,82	31.05.10	1052,68
31.01.08	1297,94	14.11.08	679,22	31.08.09	961,70	15.06.10	1070,97
15.02.08	1315,13	28.11.08	682,32	15.09.09	1021,52	30.06.10	1030,63
29.02.08	1298,11	15.12.08	678,75	30.09.09	1048,03	15.07.10	1079,16
14.03.08	1182,74	31.12.08	689,64	15.10.09	1121,69	30.07.10	1119,62
31.03.08	1183,35	15.01.09	699,94	30.10.09	1039,65	16.08.10	1108,30
15.04.08	1205,21	30.01.09	733,92	16.11.09	1127,62	31.08.10	1088,50
30.04.08	1287,40	16.02.09	790,30	30.11.09	1087,52	15.09.10	1166,50
15.05.08	1309,96	27.02.09	707,02	15.12.09	1115,20	30.09.10	1207,23
30.05.08	1298,17	16.03.09	702,93	31.12.09	1138,10	15.10.10	1267,05
13.06.08	1194,11	31.03.09	743,59	15.01.10	1155,19	28.10.10	1240,10
30.06.08	1120,62	15.04.09	826,43	29.01.10	1060,12		
15.07.08	1009,39	30.04.09	841,92	15.02.10	1053,24		
31.07.08	1073,99	15.05.09	866,74	26.02.10	1060,38		

Zdroj: Vlastní zpracování dat z www.ing.cz

Tabulka 14a - Predikční intervaly pro PF2 (Box-Jenkins)

Datum	Pořadové číslo	Bodová predikce	Intervalová predikce	Datum	Pořadové číslo	Bodová predikce	Intervalová predikce
15.11.10	74	1235,38	< 1119,4376 ; 1351,314 >	15.6.11	88	1235,12	< 807,0613 ; 1663,173 >
30.11.10	75	1219,92	< 1043,6022 ; 1396,243 >	30.6.11	89	1235,76	< 798,6876 ; 1672,823 >
15.12.10	76	1223,04	< 993,3647 ; 1452,708 >	15.7.11	90	1236,39	< 790,8957 ; 1681,883 >
31.12.10	77	1223,44	< 961,7533 ; 1485,118 >	29.7.11	91	1237,01	< 783,6262 ; 1690,385 >
14.1.11	78	1227,81	< 940,9641 ; 1514,653 >	16.8.11	92	1237,61	< 776,8395 ; 1698,384 >
31.1.11	79	1228,72	< 922,1900 ; 1535,248 >	31.8.11	93	1238,2	< 770,4910 ; 1705,912 >
15.2.11	80	1230,06	< 905,2225 ; 1554,904 >	15.9.11	94	1238,78	< 764,5473 ; 1713,008 >
28.2.11	81	1230,23	< 888,6784 ; 1571,784 >	30.9.11	95	1239,34	< 758,9756 ; 1719,704 >
15.3.11	82	1230,39	< 873,6931 ; 1588,190 >	14.10.11	96	1239,89	< 753,7490 ; 1726,029 >
31.3.11	83	1231,5	< 859,9151 ; 1603,076 >	31.10.11	97	1240,43	< 748,8417 ; 1732,008 >
15.4.11	84	1232,31	< 847,5434 ; 1617,082 >	15.11.11	98	1240,95	< 744,2312 ; 1737,665 >
29.4.11	85	1233,03	< 836,1761 ; 1629,881 >	30.11.11	99	1241,46	< 739,8968 ; 1743,021 >
16.5.11	86	1233,78	< 825,7451 ; 1641,808 >	15.12.11	100	1241,96	< 735,8195 ; 1748,097 >
31.5.11	87	1234,45	< 816,0519 ; 1652,845 >	30.12.11	101	1242,45	< 731,9820 ; 1752,909 >

Zdroj: Vlastní zpracování dat

Příloha B

Tabulka 1b – kritické hodnoty chí-kvadrát rozdělení

Stupně volnosti	$\chi^2_{0.5}$	$\chi^2_{0.25}$	$\chi^2_{0.1}$	$\chi^2_{0.05}$	$\chi^2_{0.025}$	$\chi^2_{0.01}$	$\chi^2_{0.005}$	$\chi^2_{0.001}$
1	0,5	1,3	2,7	3,8	5,0	6,6	7,9	10,8
2	1,4	2,8	4,6	6,0	7,4	9,2	10,6	13,8
3	2,4	4,1	6,3	7,8	9,3	11,3	12,8	16,3
4	3,4	5,4	7,8	9,5	11,1	13,3	14,9	18,5
5	4,4	6,6	9,2	11,1	12,8	15,1	16,7	20,5
6	5,3	7,8	10,6	12,6	14,4	16,8	18,5	22,5
7	6,3	9,0	12,0	14,1	16,0	18,5	20,3	24,3
8	7,3	10,2	13,4	15,5	17,5	20,1	22,0	26,1
9	8,3	11,4	14,7	16,9	19,0	21,7	23,6	27,9
10	9,3	12,5	16,0	18,3	20,5	23,2	25,2	29,6
11	10,3	13,7	17,3	19,7	21,9	24,7	26,8	31,3
12	11,3	14,8	18,5	21,0	23,3	26,2	28,3	32,9
13	12,3	16,0	19,8	22,4	24,7	27,7	29,8	34,5
14	13,3	17,1	21,1	23,7	26,1	29,1	31,3	36,1
15	14,3	18,2	22,3	25,0	27,5	30,6	32,8	37,7
16	15,3	19,4	23,5	26,3	28,8	32,0	34,3	39,3
17	16,3	20,5	24,8	27,6	30,2	33,4	35,7	40,8
18	17,3	21,6	26,0	28,9	31,5	34,8	37,2	42,3
19	18,3	22,7	27,2	30,1	32,9	36,2	38,6	43,8
20	19,3	23,8	28,4	31,4	34,2	37,6	40,0	45,3
21	20,3	24,9	29,6	32,7	35,5	38,9	41,4	46,8
22	21,3	26,0	30,8	33,9	36,8	40,3	42,8	48,3
23	22,3	27,1	32,0	35,2	38,1	41,6	44,2	49,7
24	23,3	28,2	33,2	36,4	39,4	43,0	45,6	51,2
25	24,3	29,3	34,4	37,7	40,6	44,3	46,9	52,6
26	25,3	30,4	35,6	38,9	41,9	45,6	48,3	54,1
27	26,3	31,5	36,7	40,1	43,2	47,0	49,6	55,5
28	27,3	32,6	37,9	41,3	44,5	48,3	51,0	56,9
29	28,3	33,7	39,1	42,6	45,7	49,6	52,3	58,3
30	29,3	34,8	40,3	43,8	47,0	50,9	53,7	59,7
40	39,3	45,6	51,8	55,8	59,3	63,7	66,8	73,4
50	49,3	56,3	63,2	67,5	71,4	76,2	79,5	86,7
60	59,3	67,0	74,4	79,1	83,3	88,4	92,0	99,6
70	69,3	77,6	85,5	90,5	95,0	100,4	104,2	112,3
80	79,3	88,1	96,6	101,9	106,6	112,3	116,3	124,8
100	99,3	109,1	118,5	124,3	129,6	135,8	140,2	149,4

Zdroj: http://www.kmt.zcu.cz/person/Kohout/info_soubory/letnisek/ruzne/tabchi.htm

Tabulka 2b - Kritické hodnoty pro Fisherův test

H	$g_t(0,01)$	H	$g_t(0,01)$	H	$g_t(0,01)$
5	0,788952	35	0,216343	80	0,107547
10	0,536014	40	0,192467	100	0,088855
15	0,407055	45	0,174027	120	0,075913
20	0,330284	50	0,159566	140	0,066393
25	0,277584	60	0,137110		
30	0,241669	70	0,120440		

Zdroj: Převzato z [6]