

UNIVERZITA PALACKÉHO V OLOMOUCI  
PŘÍRODOVĚDECKÁ FAKULTA  
KATEDRA MATEMATICKÉ ANALÝZY A APLIKACÍ MATEMATIKY

## DIPLOMOVÁ PRÁCA

Štatistická analýza časového radu cien komodít  
na burze

Vedúci diplomovej práce:  
**prof. RNDr. Ing. Lubomír Kubáček, DrSc., Dr. h. c.**  
Rok odovzdania: 2011

Vypracovala:  
**Nina Mikolajová**  
AME, II. ročník

## **Prehlásenie**

Prehlasujem, že som vytvorila túto diplomovú prácu samostatne pod vedením prof. RNDr. Ing. Lubomíra Kubáčka, DrSc., Dr. h. c. a že som v zozname použitej literatúry uviedla všetky zdroje použité pri spracovaní práce.

V Olomouci 31.3.2011

## **Pod'akovanie**

Rada by som na tomto mieste pod'akovala vedúcemu diplomovej práce prof. RNDr. Ing. Lubomírovi Kubáčkovi, DrSc., Dr. h. c. za jeho obetavú spoluprácu a čas, ktorý mi venoval. A taktiež mojej rodine, ktorá ma po celú dobu podporovala.

# Obsah

Úvod	5
<b>1 Použité značenie</b>	<b>7</b>
<b>Ekonomická časť</b>	<b>8</b>
<b>2 Vznik a vývoj búrz</b>	<b>8</b>
2.1 Najznámejšie burzy súčasnosti . . . . .	9
<b>3 Komodity a obchodovanie s nimi</b>	<b>11</b>
3.1 Finančné deriváty . . . . .	11
<b>4 Metódy odhadu ceny na burze</b>	<b>14</b>
4.1 Fundamentálna analýza . . . . .	14
4.2 Technická analýza . . . . .	15
4.3 Psychologická analýza . . . . .	16
4.4 Štatistická analýza . . . . .	16
<b>Matematická časť</b>	<b>17</b>
<b>5 Lineárny regresný model</b>	<b>17</b>
5.1 Metóda najmenších štvorcov . . . . .	17
<b>6 Časové rady</b>	<b>19</b>
6.1 Trendová zložka . . . . .	20
6.1.1 Konštantný trend . . . . .	20
6.1.2 Lineárny trend . . . . .	20
6.1.3 Kvadratický trend . . . . .	21
6.1.4 Polynomická regresia . . . . .	22
6.2 Periodická zložka . . . . .	25
6.2.1 Periodogram . . . . .	25
6.2.2 Test R. A. Fishera . . . . .	28
6.2.3 Metóda skrytých períód . . . . .	30
6.3 Kritéria posúdenia vhodnosti modelu . . . . .	31
<b>7 Rozdelenia pravdepodobností</b>	<b>33</b>
7.1 Normálne rozdelenie . . . . .	33
7.2 Studentovo rozdelenie . . . . .	35
7.3 F-rozdelenie . . . . .	36
7.4 Rozdelenie kvadratických foriem . . . . .	36

<b>8</b>	<b>Intervaly spoľahlivosti a testovanie hypotéz</b>	<b>37</b>
8.1	Intervaly spoľahlivosti vektoru regresných parametrov . . . . .	37
8.2	Testovanie významnosti regresných parametrov . . . . .	38
8.3	Jednovýberový t-test . . . . .	39
<b>9</b>	<b>Predikcie</b>	<b>40</b>
9.1	Bodové predpovede . . . . .	40
9.2	Intervalové predpovede . . . . .	40
	<b>Praktická časť</b>	<b>42</b>
<b>10</b>	<b>Štatistická analýza</b>	<b>42</b>
10.1	Časový rad ceny bavlny na burze . . . . .	42
10.2	Časový rad ceny kakaových bôbov na burze . . . . .	60
	<b>Záver</b>	<b>72</b>
	<b>Prílohy</b>	<b>74</b>
A	Trendové funkcie časového radu ceny bavlny	74
B	Trendové funkcie časového radu ceny kakaových bôbov	75
C	Fisherov test aplikovaný na hodnoty periodogramu ceny kakaových bôbov - hodnoty	76
D	Hodnoty parametrov $\delta_j, j = 1, \dots, 4$ a $\beta_i, i = 1, 2, \dots, 39$ modelu popisujúceho vývoj ceny kakaových bôbov	77
E	Priložené CD	78

# Úvod

*Neexistuje jediná oblasť matematiky, a to akokoľvek abstraktná, ktorá by sa nedala aplikovať na javy reálneho sveta.*

*N. I. Lobačevskij*

V dnešnom svete sa každý človek snaží o čo najefektívnejšie zhodnotenie svojich financií a investícia do komodít sa javí ako veľmi dobrý prostriedok. Správny obchodník však neinvestuje neuvážene, ale najprv si preskúma minulý vývoj ceny a až na základe neho môže pristúpiť k investíciám do budúcnosti. Každý investor sa zameriava na iné podnety a skúma komoditu z iného hľadiska. Preto existuje aj množstvo postupov - analýz, ktoré investorom pomáhajú. Domnievame sa, že aj pomocou matematických postupov sa dá dosiahnuť solídnych výsledkov a preto jeden z nich vyskúšame.

Cieľom práce je vykonať štatistickú analýzu časového radu cien komodít na burze. Predstaviť si základné východiská, najmä teóriu časových radov a regresnú analýzu a pomocou nich zistiť, či model obsahuje skryté periodicity a rozanalyzovať jednotlivé zložky časového radu. Použiť regresný model a určiť predikcie v časovom rade aj s konfidenčnými intervalmi. Porovnať dosiahnuté výsledky so skutočnosťou a celkový výsledok zhodnotiť.

Práca je rozdelená do troch častí. Prvá, ekonomická časť, sa zaoberá burzou a burzovými obchodmi. Prvá kapitola pojednáva o histórii a súčasnosti búr, druhá je venovaná komoditám a finančným derivátom, ktoré sa k ich obchodovaniu používajú, a tretia kapitola sa zaoberá metódami odhadu ceny na burze, ako je napríklad fundamentálna či technická analýza.

Druhá, matematická časť, vymedzuje základné teoretické východiská pre riešenie stanoveného problému - štatistickej analýzy. Obsahuje päť kapitol, ktoré sa postupne venujú lineárnemu regresnému modelu a metóde najmenších štvorcov, časovým radom a ich jednotlivým zložkám, rozdeleniam pravdepodobností, ktoré následne využijeme pri testovaní hypotéz a zostavovaní intervalov spoľahlivosti a posledná kapitola je venovaná konštrukcii bodových predpovedí a konfidenčných

intervalov.

Posledná, praktická časť, je využitím poznatkov z teoretickej časti na konkrétnom príklade štatistickej analýzy ceny bavlny a kakaových bôbov od marca 1985 do decembra 2009. Jedná sa o analýzu časového radu, jeho zložiek, zostavenie modelu a predikciu na najbližších 6 mesiacov.

# 1 Použité značenie

$\mathbb{R}^n$	$n$ -rozmerný reálny Euklidovský priestor
$\mathbf{Y}$	vektor náhodných veličín
$\mathbf{X}$	matica plánu
$\boldsymbol{\delta}, \boldsymbol{\beta}$	vektory regresných parametrov
$\hat{\boldsymbol{\delta}}, \hat{\boldsymbol{\beta}}$	odhady vektorov regresných parametrov
$\mathbf{e}, \boldsymbol{\epsilon}$	náhodný vektor
$\mathbf{I}_n$	jednotková matica rozmeru $n \times n$
$\mathbf{1}_n$	jednotkový vektor dimenzie $n$
$\mathbf{0}$	nulový vektor alebo matica
$\sigma^2$	jednotková disperzia veličiny
$\mu$	stredná hodnota veličiny
$\mathbf{A}^T$	transponovaná matica alebo vektor
$\mathbf{A}^{-1}$	inverzná matica
$\alpha$	hladina významnosti, riziko testu
$H_0, H_1(H_a)$	nulová hypotéza, alternatívna hypotéza testu
$\phi_i(\cdot)$	Čebyševov polynóm
$I(\omega)$	hodnoty periodogramu
$\omega$	frekvencia
$T$	perióda
$\mathbf{Y} \sim N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$	$n$ -rozmerný vektor má normálne rozdelenie so strednou hodnotou $\boldsymbol{\mu}$ a kovariančnou maticou $\boldsymbol{\Sigma}$
$WN(0, \sigma^2)$	biely šum (white noise)



# Ekonomická časť

## 2 Vznik a vývoj búrz

Vznik a následný vývoj búrz úzko súvisí s vývojom tržného obchodu. Trhy boli spočiatku nepravidelné verejné zhromaždenia kupujúcich a predávajúcich, ktoré sa neskôr stali pravidelnými a viazali sa najmä na cirkevné slávnosti. Burzy ako špeciálne druhy trhu sa formujú, až keď sa na trhu začína objavovať zastupiteľný tovar, čo boli v počiatkoch cenné papiere - zmenky.

V Európe začali burzy vznikať v 12. až 13. storočí najmä v talianskych mestách Janov, Benátky, Florencia a Lucca. Pojem burza sa však začal používať až v belgických Bruggách, kde sa schádzali kupci na obchodných schôdkach nazývaných "de beurse", podľa domu rodiny Van de Beurse. Stále sa však obchodovalo hlavne s mincami a zmenkami. Neskôr začali vznikať nové burzy ako napríklad burza v Lyone, Hamburgu či Londýne a rozvinuli sa aj nové druhy cenných papierov - akcie, obligácie atď., až burzy získali dnešnú podobu.

Burzy na tovar, resp. komoditné burzy sa začali rozvíjať omnoho neskôr ako burzy cenných papierov. Jedným z hlavných dôvodov boli najmä technické problémy, ktoré neumožňovali rýchle presuny veľkých množstiev tovaru. V roku 1617 bola založená jedna z prvých búrz na tovar, obilná burza v Amsterdame. Obchodníci s obilím sa tu stretávali dvakrát do týždňa a obchodovanie prebiehalo na základe vzorky. Po uzatvorení obchodu sa kupujúci šiel presvedčiť do skladu, že tovar naozaj zodpovedá vzorke.

Na začiatku sa obchodovalo iba so skutočným tovarom, ktorí bol okamžite k dispozícii. Takéto obchody sa nazývajú promptné, miestne obchody. Až koncom 19. storočia s nástupom rozvoja dopravy došlo aj k rozvoju obchodov dodacích, teda obchodov, kde tovar nie je na mieste k dispozícii. Z týchto sa neskôr vyvinuli termínové špekulačné obchody, ktoré však boli väčšinou zo strany štátu potlačované a zakazované. Najväčší rozmach zaznamenali termínové obchody v 20. stor. a to hlavne na veľkých medzinárodných burzách v USA, Veľkej Británii či v Kanade.[7]

## 2.1 Najznámejšie burzy súčasnosti

- **CME Globex**

CME Globex je najznámejšou svetovou burzou. Chicago Merkantile Exchange (CME) vznikla v roku 1898. V roku 2007 po kúpe najstaršej burzy na svete CBOT (Chicago Board of Trade) vznikla CME Group. V auguste 2008 sa pod tým istým názvom spojila aj s newyorskou burzou NYMEX. Na tejto burze je možné obchodovať prakticky čokoľvek - futures a opcie na základe úrokovej sadzby, zahraničné meny, akciové indexy, energie, poľnohospodárske komodity, kovy, ale aj alternatívne investičné produkty, ako je počasie a nehnuteľnosti. CME Globex platforma ponúka obchodovanie prakticky 24 hodín denne počas celého obchodného týždňa a možnosť obchodovať na CME majú obchodníci z viac než 80 krajín sveta. [19]

- **NYSE Euronext Liffe**

Počiatky New York Stock Exchange sa pripisujú roku 1792, kedy 24 Newyorskí burzovní makléri a obchodníci podpísali Buttonwoodsku dohodu, čím otriasli dovedy stálymi záväzkami investorov a emitentov. V roku 2006 sa NYSE spojila s Archipelago a vytvorili NYSE Group.

Burza Euronext vznikla v roku 2000 spojením Amsterdam Exchange, Brussels exchange a Paris Bourse. V roku 2002 prevzala londýnsku LIFFE (London International Financial Futures and Options Exchange). Koncom roka 2007 sa Euronext spojila s NYSE group a vznikla NYSE Euronext, čím sa spojil Európsky a Americký trh. Oficiálne ústredie sa nachádza v New Yorku, avšak európske operácie sú riadené z Paríža. Euronext ponúka široké spektrum produktov ako sú napr. futures a opcie na krátkodobé úrokové sadzby, swapy, dlhopisy, akcie či komodity. [20]

- **ICE - InterContinental Exchange**

InterContinental Exchange bola založená v máji 2000 akcionármi reprezentujúcimi niektoré najväčšie energetické spoločnosti a svetové banky, v Atlante. Misiou ICE bolo transformovať OTC trh s energiou tým, že poskytne otvorený, dostupný a 24-hodinový elektronický trh s energiou ako alternatívu k doposiaľ neprehľadnému a roztrieštenému trhu. Postupne, vďaka cenovej transparentnosti, vyššej účinnosti, likvidite a nižším nákladom oproti manuálnemu obchodovaniu, ICE vyvinula dnes vedúce elektronické trhovisko energetických komodít. V júni 2001 rozšírila svoju činnosť aj na termínových trhoch a to prvou z mnohých akvizícií - International Petroleum exchange (IPE), ktorá je momentálne Európskou vedúcou energetickou burzou futures. V roku 2005 sa ICE Futures Europe stala prvou plne elektronizovanou energetickou burzou, pričom po prvýkrát v histórii otvorila tieto trhy obchodníkom z celého sveta. ICE postupom času získala aj ďalšie trhy a indexy ako napríklad NYBOT (futures kontrakty na mäkké komodity), Russell Investments, NGI a NGX, The Clearing Corporation, Climate Exchange plc a mnoho ďalších. Dnes je platforma ICE jednou z najflexibilnejších, najefektívnejších a najbezpečnejších a slúži zákazníkom vo viac ako 70 krajinách sveta. [21]

### 3 Komodity a obchodovanie s nimi

Komodity sú základnými surovinami všetkého, čo denne jeme a používame. Sú súčasťou našej dennej rutiny, od rannej kávy alebo čaju, až po pohonné hmoty do našich áut. Môžeme ich rozdeliť do rôznych skupín napríklad :

- Obilniny - kukurica, pšenica, sójové bôby, ryža
- Mäso - hovädzie, bravčové, hydina
- Priemyselné kovy - zinok, hliník, meď
- Drahé kovy - zlato, striebro, platina
- Energie - ropa, zemný plyn, vykurovací olej
- Potraviny a vlákna - kakao, káva, bavlna, guma

Novým trendom v oblasti komodít je obchodovanie “exotických“ produktov ako je napr. prenosová kapacita telekomunikačných sietí, plynu a elektriny, deriváty počasia, deriváty emisií.

Obchodovanie komodít môže prebiehať buď priamo (promptné obchody), alebo nepriamo (termínové obchody) prostredníctvom finančných derivátov. [8] [22]

#### 3.1 Finančné deriváty

Finančné deriváty sú produkty, ktoré sa vzťahujú k určitému podkladovému aktívu alebo finančnému nástroju a ich hodnota odvodená (derivovaná) od hodnoty tohto základného finančného inštrumentu. Podkladovým aktívom finančných derivátov môžu byť akcie, dlhopisy, meny, komodity, burzové indexy. Finančné deriváty sú produkty, ktoré nám umožňujú v danom časovom okamihu zafixovať cenu alebo kurz, za ktorý sa k určitému dopredu dohodnutému budúceму dátumu môže dané aktívum kúpiť alebo predať. Tým pádom nám pomáhajú sa zaistiť proti riziku nepriaznivej zmeny ceny. Ak však vývoj nebude prebiehať

podľa našich odhadov, môžeme aj stratiť nemalý obnos finančných prostriedkov. Ak napríklad obchodník uzavrie futures kontrakt na nákup dlhopisov k určitému dátumu v cene \$10 a cena k dátumu na promptnom trhu \$5, bolo by pre obchodníka výhodnejšie nakúpiť na promptnom trhu. On však uzavrel futures kontrakt a tieto dlhopisy musí odkúpiť aj keď je to nevýhodné.

Finančné deriváty môžeme podľa vzájomného postavenia účastníkov termínového obchodu deliť na :

- *pevné (nepodmienené) deriváty*, ktoré predstavujú termínový kontrakt, pri ktorom ani jeden z partnerov nemá možnosť voľby, či daný obchod uskutoční alebo nie. Obaja účastníci sú povinní ho uskutočniť k dátumu splatnosti.
- *opčné (podmienené) deriváty* predstavujú termínový obchod, pri ktorom má majiteľ opcie právo od kontraktu odstúpiť. Jeho rozhodnutie je podmienené skutočnou cenou podkladového aktíva k dátumu splatnosti. Za túto výhodu však musí pri uzatváraní termínového kontraktu účastník v aktívnom postavení (s možnosťou voľby) zaplatiť účastníkovi v pasívnom postavení opčnú prémiiu.

Medzi najznámejšie finančné deriváty patria forwardy, futures a opcie.

## **Forwardy**

Forward je zmluva medzi dvoma stranami uzavretá v súčasnosti o povinnosti predať alebo kúpiť určité podkladové aktívum k stanovenému času v budúcnosti za cenu stanovenú v súčasnosti, pričom obchod sa uskutočňuje mimo burzy (OTC - over the counter). Všetky náležitosti forwardového kontraktu, teda cena, termín plnenia, objem obchodu, a.i. závisia iba na dohode zmluvných partnerov. Výhodou takéhoto kontraktu je úplná voľnosť pri rokovaní, určovaní podmienok obchodu. Nevýhodou však, že nakoľko forwardy nie sú obchodované na burze, predávajúci musí sám nájsť dôveryhodného záujemcu o ponúknuté finančné aktívum alebo komoditu.

## Futures

Futures je záväzná zmluva medzi dvoma stranami uzavretá v súčasnosti o povinnosti kúpiť alebo predať určité podkladové aktívum k stanovenému času v budúcnosti za cenu stanovenú v súčasnosti, pričom obchod sa uskutočňuje v štandardizovanej forme na burze. Futures obchod je tzv. pevný obchod kde sa obe strany podpisom zmluvy zaväzujú kontrakt uskutočniť. Môžeme povedať, že futures je štandardizovaný forward, teda forward obchodovaný na burze, pričom táto na seba preberá záruku serióznosti.

## Opcie

Názov opcia pochádza z anglického option, čiže voľba. Vlastník opcie si teda môže zvoliť, či zjednaný obchod zrealizuje alebo nie. Za toto právo však musí zaplatiť opčnú prémiiu. Napríklad ak obchodník vlastní opciu na nákup akcií spoločnosti KFC za \$100 a spoločnosť predáva svoje akcie za \$120, musela by byť opčná prémia rovná aspoň \$20. Vtedy by obchodník opciu nevyužil a nakúpil by akcie radšej na voľnom trhu.

Na trhu existujú dva základné druhy opcií a to call a put. Call opcia dáva vlastníčkovi právo kúpiť určitý tovar za zjednanú cenu k určitému dátumu. Put opcia naopak dáva vlastníčkovi právo predať dohodnutý tovar k určitému dátumu. Podľa termínu plnenia rozoznávame dva typy opcií. Európske opcie, ktoré sú splatné iba k dátumu splatnosti a americké opcie, pri ktorých môže plnenie nastať kedykoľvek v dobe do dátumu splatnosti.

V súčasnosti sú najznámejšie opcie na akcie a dlhopisy, zahraničné meny, akciové indexi, futures kontrakty.

[11] [16]

## 4 Metódy odhadu ceny na burze

Každý obchodník na burze by najradšej vedel presný vývoj cien všetkých obchodovateľných inštrumentov. Vtedy by sa ľahko rozhodol, ktoré inštrumenty do svojho portfólia zaradiť, ktoré nie, na ako dlho a aký zisk má očakávať. Tieto informácie však nemá a preto vznikajú rôzne metódy ako vývoj cien predpovedať. Medzi najznámejšie patria technická analýza, fundamentálna analýza, psychologická analýza, štatistická analýza.

### 4.1 Fundamentálna analýza

Fundamentálna analýza je komplexný a veľmi rozsiahly proces, pri ktorom obchodník sleduje a vyhľadáva faktory, ktoré ovplyvňujú ním obchodovaný finančný inštrument. Záber fundamentálnej analýzy je veľmi široký - od ceny inštrumentu až po globálne vplyvy, ktoré by mohli určitým spôsobom ovplyvniť vývoj na danom trhu.

Fundamentálna analýza prebieha na troch úrovňach:

1. **Makroekonomická (globálna) fundamentálna analýza**  
skúma celkovú ekonomickú situáciu (úrokové sadzby, inflácia, HDP)
2. **Odvetvová (sektorová) analýza**  
skúma jednotlivé odvetvia a ich špecifiká (ťažobný, stavebný, potravinový sektor)
3. **Analýza konkrétneho finančného inštrumentu**  
akcie, komodity, menového páru a podobne

[17] [23]

## 4.2 Technická analýza

Technická analýza je proces štúdia trhov, ktorý sa zameriava výhradne na grafy a rôzne pomocné indikátory, ktoré sa v grafoch dajú identifikovať, za účelom predpovedania trendov budúceho vývoja. Vychádza z troch základných predpokladov :

### 1. Ceny zohľadňujú všetko

Technickí analytici veria, že všetko čo môže ovplyvniť cenu (fundamentálne, psychologické, politické vplyvy), je už v cene obsiahnuté. Nie je teda nutné sledovať žiadne iné informácie, iba cenové pohyby. Vedia, že ak dopyt prevyšuje ponuku, ceny rastú a opačne, ale nezameriavajú sa na dôvody, prečo tento stav nastáva.

### 2. Ceny sa pohybujú v trendoch

Tento predpoklad je pre technickú analýzu kľúčový. Účelom štúdia grafov je identifikácia trendu v rannom štádiu, aby investor potom mohol obchodovať v smere trendu a platí, že trend ostáva v platnosti až kým nedôjde k jeho obratu. Tento jav je nutné včas odhaliť a tak sa neocitnúť v stratovej pozícii. Na každom cenovom grafe je možné nájsť rôzne druhy trendov (krátkodobé, strednedobé, dlhodobé) a závisí iba na investovi, ktorý trend bude sledovať, podľa toho, na aký dlhý časový okamih chce investovať.

### 3. História sa opakuje

Táto skutočnosť vychádza najmä z ľudskej psychiky, ktorá sa rokmi nemení. Na cenových grafoch boli identifikované určité vzory správania a predpokladá sa, že keď tieto platili v minulosti, budú rovnako dobre platiť aj v budúcnosti. Kľúčom k predpovedaniu budúcnosti je teda štúdium minulosti a budúcnosť bude iba zopakovaná minulosť.

Veľkou výhodou je, nakoľko sa technická analýza dá aplikovať na každý trh – akcie, dlhopisy, komoditné futures, meny a i., že si analytik môže vždy vybrať a obchodovať trh, na ktorom sa práve niečo deje a nie je ako fundamentalista viazaný na jeden konkrétny.[15]



### 4.3 Psychologická analýza

Ako už názov napovedá, psychologický analytici sa pri skúmaní cien finančných inštrumentov zameriavajú na psychologické chovanie investorov. Vychádzajú z presvedčenia, že pohyb kurzu je iba následok správania investorov, pričom základom takéhoto chovania je ľudská psychika. Psychologickí analytici hľadajú impulzy, ktoré vedú investorov k masovejším nákupom resp. predajom, ktoré následne spôsobujú pohyby kurzu smerom nahor alebo nadol.

Predmetom skúmania psychologickéj analýzy teda nie je finančný inštrument, ale človek a impulzy na neho pôsobiace. Preto sa psychologická analýza zaoberá napr. skúmaním psychológie davu, ktorú veľmi precízne vysvetlil Gustave Le Bon. Ďalšie známe investičné psychologické analytické koncepcie sú napr. prístupy André Kostolanyho, Johna Maynarda Keynesa, Georga Drasnara, či teória špekulatívnych bublín a iné, o ktorých je možné sa bližšie dočítať v [18] str. 90 - 94. [18]

### 4.4 Štatistická analýza

Pri štatistickej analýze vývoja cien na burze sa využívajú mnohé štatistické postupy ako napr. exponenciálne vyrovnávanie, teória časových radov, Box - Jenkinsonova metodológia a iné. Tieto metódy sú veľmi rozsiahle preto si v nasledujúcich kapitolách predstavíme iba jednu - teóriu časových radov (dekompozičný prístup). Najprv sa však musíme zoznámiť s lineárnym regresným modelom.

# Matematická časť

## 5 Lineárny regresný model

Predpokladajme, že máme náhodné veličiny  $Y_1, Y_2, \dots, Y_n$  a maticu  $\mathbf{X} = (x_{ij})$  typu  $n \times k$ , pričom  $k < n$ , takže máme viac pozorovaní ako parametrov. Matica  $\mathbf{X}$  je matica konštánt, nazývajúca sa tiež matica plánu. Ďalej predpokladajme, že pre náhodný vektor  $\mathbf{Y} = (Y_1, Y_2, \dots, Y_n)^T$  platí

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e},$$

pričom  $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_k)^T$  je vektor neznámych parametrov a  $\mathbf{e} = (e_1, e_2, \dots, e_n)^T$  je náhodný vektor, ktorý splňuje podmienky  $E\mathbf{e} = 0$  a  $\text{var}\mathbf{e} = \sigma^2\mathbf{I}$ . Parameter  $\sigma^2 > 0$  je taktiež neznámym parametrom.

Vektor  $\mathbf{X}\boldsymbol{\beta}$  je nenáhodný a platí  $E\mathbf{Y} = \mathbf{X}\boldsymbol{\beta}$  a  $\text{var}\mathbf{Y} = \sigma^2\mathbf{I}$ . Takýto model sa nazýva regresný model, presnejšie lineárny regresný model, nakoľko  $\mathbf{Y}$  závisí na  $\boldsymbol{\beta}$  lineárne.

Pri konštrukcii modelu predpokladáme, že nebudeme zavádzať nadbytočné vysvetľujúce premenné  $x_{ij}$  teda požadujeme, aby matica  $\mathbf{X}$  mala lineárne nezávislé stĺpce. Nakoľko predpokladáme  $k < n$ , dostávame  $h(\mathbf{X}) = k$ . Navyše z daného predpokladu platí, že matica  $\mathbf{X}^T\mathbf{X}$  je regulárna.[3][13]

### 5.1 Metóda najmenších štvorcov

Na odhad parametra  $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_k)^T$  v lineárnom regresnom modeli použijeme najpoužívanejšiu metódu, metódu najmenších štvorcov. Pri tejto metóde minimalizujeme výraz  $(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^T(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})$  ako funkciu parametra  $\boldsymbol{\beta}$ . Výsledný minimálny odhad označme  $\hat{\boldsymbol{\beta}} = (\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_k)^T$ .

**Veta 5.1.** *Metódou najmenších štvorcov dostávame v lineárnom regresnom modeli odhad*

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}$$

**Dôkaz 5.1.** Ako sme už uviedli pri metóde najmenších štvorcov minimalizujeme výraz

$$(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^T(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}).$$

Nájdeme teda deriváciu daného výrazu podľa vektoru  $\boldsymbol{\beta}$

$$\begin{aligned} \frac{\partial(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^T(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})}{\partial\boldsymbol{\beta}} &= \frac{\partial(\mathbf{Y}^T\mathbf{Y} - 2\mathbf{Y}^T\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\beta}^T\mathbf{X}^T\mathbf{X}\boldsymbol{\beta})}{\partial\boldsymbol{\beta}} = \\ &= -2\mathbf{X}^T\mathbf{Y} + 2\mathbf{X}^T\mathbf{X}\boldsymbol{\beta} \stackrel{!}{=} 0 \\ \mathbf{X}^T\mathbf{X}\boldsymbol{\beta} &= \mathbf{X}^T\mathbf{Y} / * (\mathbf{X}^T\mathbf{X})^{-1} \end{aligned}$$

Matica  $(\mathbf{X}^T\mathbf{X})^{-1}$  existuje vďaka predpokladu regularity. Odhad parametra  $\boldsymbol{\beta}$  dostaneme v tvare

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}$$

Sústava lineárnych rovníc  $\mathbf{X}^T\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}^T\mathbf{Y}$ , z ktorej sa počíta vektor  $\hat{\boldsymbol{\beta}}$ , sa nazýva sústava normálnych rovníc. Vypočítaný vektor

$$\hat{\mathbf{Y}} = \mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}$$

môžeme považovať za najlepšiu aproximáciu vektora  $\mathbf{Y}$  aká sa dá vytvoriť lineárnou kombináciou stĺpcov matice  $\mathbf{X}$ .

**Veta 5.2.** Platí  $E\hat{\boldsymbol{\beta}} = \boldsymbol{\beta}$  a  $\text{var}\hat{\boldsymbol{\beta}} = \sigma^2(\mathbf{X}^T\mathbf{X})^{-1}$ .

**Dôkaz 5.2.**

$$\begin{aligned} E\hat{\boldsymbol{\beta}} &= E(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^TE\mathbf{Y} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{X}\boldsymbol{\beta} = \boldsymbol{\beta} \\ \text{var}\hat{\boldsymbol{\beta}} &= (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T(\text{var}\mathbf{Y})\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T(\sigma^2\mathbf{I})\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1} = \\ &= \sqrt{\sigma^2(\mathbf{X}^T\mathbf{X})^{-1}} \end{aligned}$$

Z predchádzajúcej vety vyplýva, že odhad  $\hat{\boldsymbol{\beta}}$  metódou najmenších štvorcov je nestranným odhadom vektoru parametrov  $\boldsymbol{\beta}$ . [3][13]

## 6 Časové rady

Časovým radom rozumieme postupnosť vecne a priestorovo zrovnateľných pozorovaní, ktoré sú z časového hľadiska zoradené jednoznačne. Najjednoduchší spôsob modelovania časového radu je jednorozmerný model symbolicky zapísaný

$$y_t = f(t), \quad t = 1, 2, \dots, n,$$

kde  $y_t$  je teoretická (modelovaná) hodnota v čase  $t$  a  $f(t)$  je modelujúca funkcia. Pri modelovaní hodnôt sa snažíme o to, aby bol rozdiel  $y_t - \hat{y}_t$  reálnej a odhadnutej hodnoty označovaný  $\epsilon_t$  v súčte pre každé  $t$  čo najmenší.

Na časový rad existuje mnoho pohľadov a prístupov, my si však vysvetlíme iba klasický (dekompozičný) prístup. Tento vychádza z dekompozície (rozkladu) časového radu na štyri zložky a to trendovú označovanú  $T_t$ , sezónnu  $S_t$ , cyklickú  $C_t$  a náhodnú  $\epsilon_t$ . Tento rozklad môže byť dvojakého tvaru:

- **aditívny tvar**

$$Y_t = T_t + S_t + C_t + \epsilon_t, \quad t = 1, 2, \dots, n$$

- **multiplikatívny tvar**

$$Y_t = T_t S_t C_t \epsilon_t, \quad t = 1, 2, \dots, n$$

Výber tvaru je podmienený skúmaným časovým radom. Vyberieme ten, ktorý lepšie modeluje dáta. Teraz si bližšie rozoberme jednotlivé zložky:

- *trendová zložka* nám určuje dlhodobú tendenciu vývoja sledovaného časového radu. Trend poznáme rastúci, klesajúci alebo konštantný, ktorý pozorujeme ak hodnoty sledovaného ukazovateľa kolísajú okolo určitej hodnoty.
- *sezónna zložka* je pravidelná odchýlka od trendovej zložky. Táto odchýlka sa objavuje s periodicitou rovnou alebo menšou ako jeden rok. Príčiny tohto kolísania sú rôzne, napr. striedanie ročných období, pracovných dní a víkendov.

- *cyklická zložka* je podobná sezónnej zložke ale ide o kolísanie okolo trendu s dĺžkou väčšou ako jeden rok. Často sa pri analýze časových radov cyklická a sezónna zložka nazýva súhrnne *periodická zložka* alebo sa cyklická zložka analyzuje ako súčasť trendovej zložky.
- *náhodná zložka* ostáva v časovom rade po eliminácii trendovej, sezónnej a cyklickej zložky. Jej zdrojom by mali byť malé a nepostihnuteľné príčiny. V tomto prípade je možné ju štatistiky vyjadriť ako biely šum  $WN(0, \sigma^2)$ .

[5][6][10]

## 6.1 Trendová zložka

Ako sme už napísali, trend určuje dlhodobú tendenciu vývoja, smerovania hodnôt sledovaného ukazovateľa. Existuje mnoho funkcií modelujúcich trendovú zložku a v nasledujúcich podkapitolách si niektoré z nich uvedieme.[10][6]

### 6.1.1 Konštantný trend

Konštantný trend je trend charakterizovaný predpisom

$$T_t = \beta_0, \quad t = 1, 2, \dots, k.$$

Parameter  $\beta_0$  je neznámy, ale ľahko odhadnuteľný pomocou metódy najmenších štvorcov.

$$\hat{\beta}_0 = (\mathbf{1}_k^T \mathbf{1}_k)^{-1} \mathbf{1}_k^T \mathbf{T}_t = \sum_{t=1}^k T_t / k = \bar{T}_t,$$

kde  $\bar{T}_t$  je priemer všetkých hodnôt časového radu a  $k$  je dĺžka časového radu,  $\mathbf{1}_k = (1, 1, \dots, 1)^T$  a  $\mathbf{T}_t = (T_1, \dots, T_k)^T$ . [5]

### 6.1.2 Lineárny trend

Lineárny trend môžeme charakterizovať rovnicou

$$T_t = \beta_0 + \beta_1 t, \quad t = 1, 2, \dots, k$$

pričom parameter  $t$  je časová premenná,  $T_t$  namerané hodnoty a hľadanými parametrami sú absolútny člen  $\beta_0$  a  $\beta_1$  nazývaný tiež smernica alebo prírastok za časovú jednotku. Najlepšie nevychýlené odhady parametrov  $\beta_0$  a  $\beta_1$  (označme  $\hat{\beta}_0$  a  $\hat{\beta}_1$ ) získame metódou najmenších štvorcov, ktorú môžeme použiť nakoľko je uvedená funkcia lineárna vzhľadom k parametrom. Označme

$$\hat{\boldsymbol{\beta}} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix} \quad \mathbf{X} = \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ \cdot & \cdot \\ \cdot & \cdot \\ 1 & k \end{pmatrix}.$$

Potom pre  $\hat{\boldsymbol{\beta}}$  platí :  $\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{T}_t$ . Po dopočítaní dostávame odhady

$$\hat{\beta}_0 = \frac{\sum t^2 \sum T_t - \sum t \sum t T_t}{n \sum t^2 - (\sum t)^2} \quad \hat{\beta}_1 = \frac{n \sum t T_t - \sum t \sum T_t}{n \sum t^2 - (\sum t)^2},$$

pričom  $\sum = \sum_{t=1}^k$ . [5][10]

### 6.1.3 Kvadratický trend

V časovom rade je kvadratický trend charakterizovaný rovnicou

$$T_t = \beta_0 + \beta_1 t + \beta_2 t^2, \quad t = 1, 2, \dots, k.$$

Neznámymi parametrami sú  $\beta_0$ ,  $\beta_1$  a  $\beta_2$ , parameter  $t$  je rovnaký ako v prechádzajúcom prípade. Matica  $\mathbf{X}$  bude tentokrát rozmeru  $k \times 3$  a bude vyzeráť

$$\mathbf{X} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 4 \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ 1 & k & k^2 \end{pmatrix}$$

Parametre v modeli odhadneme rovnako ako v predchádzajúcom prípade metódou najmenších štvorcov. Explicitné tvary týchto odhadov sú už dosť zložité a je možné ich nájsť v [10] str. 103. [5][10]

### 6.1.4 Polynomická regresia

Pri určovaní trendovej funkcie by sme mohli postupovať ďalej podobne ako v predchádzajúcich prípadoch a to zvyšovaním stupňa polynómu

$$T_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \dots + \beta_n t^n, \quad t = 1, 2, \dots, k$$

a to až na polynóm  $n$ -tého stupňa. Zvyšovaním stupňa polynómu sa zvyšuje aj hodnosť matice

$$\mathbf{X} = \begin{pmatrix} 1 & 1 & , \dots, & 1^n \\ 1 & 2 & , \dots, & 2^n \\ \cdot & \cdot & , \dots, & \cdot \\ \cdot & \cdot & , \dots, & \cdot \\ \cdot & \cdot & , \dots, & \cdot \\ 1 & k & , \dots, & k^n \end{pmatrix}$$

a v tomto prípade má vektor  $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_n)^T$  dimenziu  $n + 1$ .

Modelované hodnoty  $Y_t, t = 1, 2, \dots, k$  môžeme potom maticovo vyjadriť ako

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \quad \boldsymbol{\beta} \in \mathbb{R}^n, \text{var}(\boldsymbol{\epsilon}) = \boldsymbol{\Sigma}$$

Pre odhad parametra  $\boldsymbol{\beta}$  napríklad najbežnejšou metódou, metódou najmenších štvorcov, je nutné vypočítať maticu  $(\mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{X})^{-1}$ , čo môže byť pri väčších dimenziách vektoru  $\boldsymbol{\beta}$  nepohodlná a taktiež numericky málo stabilná. Preto je vhodné použiť pri polynomickej regresii Čebyševovu metódu.

**Definícia 6.1.** Označme  $t_1, t_2, \dots, t_n$  experimentálne body (body, v ktorých sa realizuje meranie) a kovariančná matica vektoru  $\mathbf{Y}$ , vektoru nameraných hodnôt, nech má tvar  $\text{var}(\mathbf{Y}) = \sigma^2 \mathbf{I}$ .

Čebyševovými polynómami nazveme takú  $n$ -ticu polynómov

$$\phi_i(\cdot), \quad i = 0, 1, \dots, n,$$

pre ktoré platí

$$\sum_{j=1}^n \phi_i(t_j) \phi_k(t_j) = 0,$$

práve vtedy ak  $i \neq k, i, k = 0, 1, \dots, n$ .

Pri určovaní Čebyševových polynómov  $\phi_0(\cdot), \phi_1(\cdot)$ , atď. postupujeme nasledovne:

$$\phi_0(t_j) = 1, \quad t_j \in \mathbb{R}^1$$

$$\phi_1(t_j) = t_j - c_{1,0}\phi_0(t_j),$$

kde  $\phi_1(\cdot)$  musí splňovať podmienku

$$\sum_{j=1}^n \phi_0(t_j)\phi_1(t_j) = \sum_{j=1}^n t_j\phi_0(t_j) - c_{1,0} \sum_{j=1}^n \phi_0^2(t_j) = 0.$$

Z nej po dopočítaní vyplýva, že

$$c_{1,0} = -\frac{1}{n} \sum_{j=1}^n t_j = -\bar{t}.$$

Pre  $\phi_2$  platí  $\phi_2(t_j) = t_j^2 - c_{2,1}\phi_1(t_j) - c_{2,0}\phi_0(t_j)$  a tiež musí splňať podmienky :

$$\sum_{j=1}^n \phi_2(t_j)\phi_0(t_j) = \sum_{j=1}^n t_j^2\phi_0(t_j) - c_{2,0} \sum_{j=1}^n \phi_0^2(t_j) = 0 \quad \Rightarrow \quad c_{2,0} = \frac{\sum_{j=1}^n t_j^2\phi_0(t_j)}{\sum_{j=1}^n \phi_0^2(t_j)}$$

$$\sum_{j=1}^n \phi_2(t_j)\phi_1(t_j) = \sum_{j=1}^n t_j^2\phi_1(t_j) - c_{2,1} \sum_{j=1}^n \phi_1^2(t_j) = 0 \quad \Rightarrow \quad c_{2,1} = \frac{\sum_{j=1}^n t_j^2\phi_1(t_j)}{\sum_{j=1}^n \phi_1^2(t_j)}$$

Všeobecne môžeme písať vzťahy pre  $s$ -tý polynóm :

$$\phi_s(t_j) = t_j^s - c_{s,s-1}\phi_{s-1}(t_j) - \dots - c_{s,0}\phi_0(t_j),$$

pričom

$$\sum_{j=1}^n \phi_s(t_j)\phi_k(t_j) = 0, \quad k = 0, 1, \dots, s-1.$$



Nakoľko

$$\sum_{j=1}^n \phi_s(t_j) \phi_k(t_j) = \sum_{j=1}^n t_j^s \phi_k(t_j) + c_{s,k} \sum_{j=1}^n \phi_k^2(t_j)$$

potom zrejme

$$c_{s,k} = - \frac{\sum_{j=1}^n t_j^s \phi_k(t_j)}{\sum_{j=1}^n \phi_k^2(t_j)}$$

Model experimentu teraz môžeme zapísať maticovo :

$$\mathbf{Y} = \begin{pmatrix} \phi_0(t_1) & \phi_1(t_1) & \dots & \phi_{m-1}(t_1) \\ \phi_0(t_2) & \phi_1(t_2) & \dots & \phi_{m-1}(t_2) \\ \vdots & \vdots & \dots & \vdots \\ \phi_0(t_n) & \phi_1(t_n) & \dots & \phi_{m-1}(t_n) \end{pmatrix} \begin{pmatrix} \delta_1 \\ \delta_2 \\ \vdots \\ \delta_m \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix} = \mathbf{X}\boldsymbol{\delta} + \boldsymbol{\epsilon}, \quad \text{var}(\mathbf{Y}) = \sigma^2 \mathbf{I}.$$

Potom

$$\mathbf{X}^T \mathbf{X} = \begin{pmatrix} \sum_{j=1}^n \phi_0^2(t_j) & 0 & \dots & 0 \\ 0 & \sum_{j=1}^n \phi_1^2(t_j) & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & \sum_{j=1}^n \phi_{m-1}^2(t_j) \end{pmatrix}$$

a pre odhad parametra  $\hat{\delta}_i$  dostávame vzťah :

$$\hat{\delta}_i = \frac{\sum_{j=1}^n y_j \phi_{i-1}(t_j)}{\sum_{j=1}^n \phi_{i-1}^2(t_j)}, \quad i = 1, 2, \dots, m$$

$$\text{var}(\hat{\delta}) = \sigma^2 \begin{pmatrix} 1/\sum_{j=1}^n \phi_0^2(t_j) & 0 & \dots & 0 \\ 0 & 1/\sum_{j=1}^n \phi_1^2(t_j) & \dots & 0 \\ \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \dots & \cdot \\ 0 & 0 & \dots & 1/\sum_{j=1}^n \phi_{m-1}^2(t_j) \end{pmatrix}$$

Ďalšou veľkou výhodou použitia tejto metódy je nekorelovanosť odhadov  $\hat{\delta}_1, \hat{\delta}_2, \dots, \hat{\delta}_m$ . Ak teda potrebujeme zvýšiť stupeň polynómu, nemusíme prepočítavať všetky koeficienty, ale iba dopočítame pridávané. [12]

## 6.2 Periodická zložka

### 6.2.1 Periodogram

Ako sme už spomínali, vo vývoji cien komodít na burze sa často vyskytuje určitá periodickosť, trend vývoja cien sa väčšinou s nejakým časovým odstupom opakuje. Aby sme mohli túto vlastnosť matematicky sledovať, pomôže nám štatistický postup hľadania významných frekvencií/periód nazývaný periodogram, ktorý bližšie popísal Jiří Anděl v [1].

Takmer každý časový rad môže byť vyjadrený ako súčet sínusových (cosínusových) vln s rôznou periódou (doba trvania jedného cyklu) a amplitúdou (max/min hodnoty počas cyklu). A práve tento fakt môžeme využiť pri skúmaní periodického (cyklického) správania časového radu.

**Definícia 6.2.** *Uvažujme konečnú postupnosť náhodných veličín  $Y_1, \dots, Y_n$  a funkciu  $I(\omega)$  definovanú vzorcom*

$$I(\omega) = \frac{1}{2\pi n} \left| \sum_{t=1}^n Y_t e^{-it\omega} \right|^2, \quad -\pi \leq \omega \leq \pi, \quad (1)$$

*Funkcia  $I(\omega)$  sa potom nazýva periodogramom postupnosti  $Y_1, \dots, Y_n$ .*

Značenie:

perióda  $T$  počet časových jednotiek  $t$  nutných na dokončenie jedného sínusového (cosínusového) cyklu  
 frekvencia  $\omega = 2\pi/T$  podiel celého cyklu ukončený za jednu časovú jednotku  
 Predpokladajme teraz, že

$$Y_t = \sum_{k=1}^p a_k e^{it\omega_k} + X_t, \quad t = 1, 2, \dots, n \quad (2)$$

kde  $a_1, \dots, a_k$  sú nenulové konštanty,  $\omega_1, \dots, \omega_p$  sú navzájom rôzne čísla z intervalu  $(-\pi; \pi)$  a  $\{X_t\}$  je postupnosť nekorelovaných náhodných veličín s nulovou strednou hodnotou a s rovnakým rozptylom  $\sigma^2 > 0$ . Periodická zložka  $\sum a_k e^{it\omega_k}$  je teda prekrytá postupnosťou náhodných veličín  $\{X_t\}$ .

Za platnosti (2) môžeme po dosadení a pre násobení písať :

$$n^{-1/2} \sum_{t=1}^n Y_t e^{-it\omega} = n^{-1/2} \sum_{k=1}^p a_k \sum_{t=1}^n e^{it(\omega_k - \omega)} + n^{-1/2} \sum_{t=1}^n X_t e^{-it\omega} \quad (3)$$

**Veta 6.1.** Ak je  $\omega$  rôzne od všetkých  $\omega_1, \dots, \omega_p$ , potom je prvý člen na pravej strane vzorca (3) rovný

$$n^{-1/2} \sum_{k=1}^p a_k e^{i(\omega_k - \omega)} \frac{e^{in(\omega_k - \omega)} - 1}{e^{i(\omega_k - \omega)} - 1}$$

Pre  $n \rightarrow \infty$  tento člen konverguje k nule. Ak sa  $\omega = \omega_j$ , potom je prvý člen rovný

$$n^{1/2} a_j + n^{-1/2} \sum_{\substack{k=1 \\ k \neq j}}^p a_k e^{i(\omega_k - \omega_j)} \frac{e^{in(\omega_k - \omega_j)} - 1}{e^{i(\omega_k - \omega_j)} - 1}$$

a pri  $n \rightarrow \infty$  narastá jeho absolútna hodnota nad všetky medze.

**Dôkaz 6.1.** Vzorce dostaneme jednoduchým dosadením do vzorca (3) pre  $\omega = \omega_j$  a  $\omega \neq \omega_j$ , pre  $\forall j$ . Ide totiž o súčet geometrickej postupnosti. V prvom prípade pre  $\omega \neq \omega_j$ ,  $\forall j$  a  $n \rightarrow \infty$  platí

$$\lim_{n \rightarrow \infty} n^{-1/2} \sum_{k=1}^p a_k e^{i(\omega_k - \omega)} \frac{e^{in(\omega_k - \omega)} - 1}{e^{i(\omega_k - \omega)} - 1} = 0$$

Pre  $\omega = \omega_j, \forall j$  môžeme prvý člen výrazu pre  $n \rightarrow \infty$  vypočítať ako

$$\lim_{n \rightarrow \infty} n^{1/2} a_j = \infty,$$

pre druhý člen platí podobný vzťah ako v prípade  $\omega \neq \omega_j, \forall j$  a jeho limita pre  $n \rightarrow \infty$  je rovná nule, takže limita celého výrazu je rovná  $\infty$ .

Druhý člen na pravej strane vzorca (3) je náhodná veličina s nulovou strednou hodnotou, ktorá má rozptyl

$$E \frac{1}{n} \left( \sum_{t=1}^n X_t e^{-it\omega} \right)^2 = \frac{1}{n} \sum_{t=1}^n \text{var} X_t = \sigma^2.$$

Rozptyl na  $n$  nezávisí, môžeme preto konštatovať, že za platnosti modelu (2) bude mať pre veľké  $n$  periodogram výrazne veľké hodnoty (rádu  $n$ ) v bodoch  $\omega_1, \dots, \omega_p$ . Ostatné hodnoty budú relatívne malé. Budú kolísať približne okolo  $\sigma^2/2\pi$ . Pre jednoduchšie a lepšie vyšetovanie štatistických vlastností periodogramu si odvodíme iný tvar vzorca (1).

**Lemma 6.1.** *Položme teraz*

$$A(\omega) = (2/n)^{1/2} \sum_{t=1}^n Y_t \cos t\omega, \quad B(\omega) = (2/n)^{1/2} \sum_{t=1}^n Y_t \sin t\omega.$$

*Potom platí*

$$I(\omega) = \frac{1}{4\pi} A^2(\omega) + \frac{1}{4\pi} B^2(\omega).$$

**Dôkaz 6.2.**

$$\begin{aligned} I(\omega) &= \frac{1}{2\pi n} \left| \sum_{t=1}^n Y_t e^{-it\omega} \right|^2 = \frac{1}{2\pi n} \left| \sum_{t=1}^n Y_t \cos t\omega - i \sum_{t=1}^n Y_t \sin t\omega \right|^2 = \\ &= \frac{1}{2\pi n} \left[ \left( \sum_{t=1}^n Y_t \cos t\omega \right)^2 + \left( \sum_{t=1}^n Y_t \sin t\omega \right)^2 \right] = \frac{1}{4\pi} [A^2(\omega) + B^2(\omega)]. \end{aligned}$$

Najkratšia zistiteľná perióda má dĺžku  $T = 2$ . Odpovedá jej frekvencia  $\omega = \pi$ , ktorá sa nazýva Nyquistova frekvencia. [1][5][9]

### 6.2.2 Test R. A. Fishera

Test R. A. Fishera používame na zisťovanie významnosti najvyšších hodnôt periodogramu. Uvažujme postupnosť nezávislých náhodných veličín  $Y_1, \dots, Y_n$ . Testovaná hypotéza je:

$$H_0 : Y_t = \epsilon_t \quad \epsilon_t \sim WN(0, \sigma^2)$$

Predpokladajme, že počet pozorovaní  $n = 2m + 1$ , čiže  $n$  je nepárne číslo. (Ak by  $n$  bolo párne, obvykle sa vynechá prvý člen, takže člen časovo najvzdialenejší od súčasnosti). Hodnoty periodogramu

$$I(\omega_k), \quad k = 1, \dots, m.$$

v bodoch

$$\omega_k = \frac{2\pi}{n}k \quad k = 1, \dots, m.$$

zoradíme podľa veľkosti

$$V_1 \geq V_2 \geq \dots \geq V_m.$$

Položme (tzv. Fisherova štatistika)

$$W = \frac{V_1}{V_1 + \dots + V_m},$$

táto nadobúda hodnoty medzi 0 a 1. Ak by boli všetky hodnoty skoro rovnaké, bude hodnota  $W$  blízka číslu  $\frac{1}{m}$ . Naopak, ak by veličina  $V_1$  nadobúdala výrazne vyššiu hodnotu v porovnaní s ostatnými veličinami  $V_2, \dots, V_m$ , bude hodnota  $W$  blízka číslu 1.

Je teda vidieť, že veľké hodnoty, tj. také, ktoré sú blízko 1, budú tvoriť kritický obor testovanej hypotézy proti alternatíve :

$$H_1 : Y_t = (\alpha \cos t\omega + \beta \sin t\omega) + \epsilon_t, \quad \epsilon_t \sim WN(0, \sigma^2)$$

Fisher odvodil distribučnú funkciu štatistiky  $W$  za platnosti hypotézy  $H_0$  (za predpokladu uvažovania gaussovského bieleho šumu):

$$1 - F_{W|H_0}(x) = P(W > x|H_0) = m(1-x)^{m-1} - \binom{m}{2}(1-2x)^{m-1} \dots,$$

kde  $0 < x < 1$  a na pravej strane sčítame tak dlho, kým sú členy  $(1 - kx)$  kladné. Čo sa dá taktiež súhrnne zapísať ako:

$$P(W > x|H_0) = \sum_{k=1,2,\dots} \binom{m}{k} [\max(0, 1 - kx)]^{m-1} = \sum_{k=1,2,\dots} \binom{m}{k} [(1 - kx)_+]^{m-1} \quad (4)$$

Už R. A. Fisher zistil, že pre  $m \leq 50$  stačí použiť prvým členom na pravej strane rovnosti 4, teda pre  $k = 1$ . Tabuľka ukazujúca túto skutočnosť je uverejnená v [1] str. 82. Pri nižších hladinách významnosti ako 5% je dokonca aproximácia ešte lepšia.

Hypotézu  $H_0$  na hladine významnosti  $\alpha$  zamietame ak:

$$1 - F_{W|H_0}(\omega) = P(W > \omega|H_0) = \alpha_{W|H_0} \leq \alpha,$$

kde  $\omega$  je skutočná hodnota Fisherovej štatistiky pri daných hodnotách skutočného časového radu a  $\alpha_{W|H_0}$  je tzv. p-value (p-hodnota testu).

Po otestovaní prvej hodnoty  $V_1$  a zistení, že ide o významnú (významnú) periodicitu, pri testovaní ďalšej periodicity hodnotu  $V_1$  vynechávame. Teda ďalej na základe veličín  $V_2, \dots, V_m$  položíme

$$W^{(2)} = \frac{V_2}{V_2 + \dots + V_m}$$

a stanovíme  $P(W^{(2)} > \omega^{(2)})$  podľa rovnakého vzorca, kde namiesto  $m$  dosadíme  $m - 1$ . Ak vyjde aj druhá hodnota významná, opäť sa táto hodnota vynechá a  $m$  sa zmenší o 1. Keď takto získame frekvencie  $\omega_1, \dots, \omega_p$ , získame model so

známymi frekvenciami a neznáme parametre odhadneme metódou najmenších štvorcov.[1][9]

### 6.2.3 Metóda skrytých periód

Hlavným predpokladom pre použitie metódy skrytých periód je fakt, že časový rad je očistený od trendu. Takto očistený časový rad môžeme symbolicky zapísať

$$y'_t = y_t - \hat{T}_t, \quad t = 1, 2, \dots, n$$

kde  $y'_t$  je očistený časový rad,  $y_t$  je pôvodný časový rad, ktorý skúmame a  $\hat{T}_t$  je odhad trendovej funkcie.

Uvažujme očistené časové rady, ktoré môžeme rozložiť na súčet harmonických frekvencií, pre ktoré sa dĺžky periód dajú vyjadriť ako podiel  $T_k = \frac{n}{k}$ , pričom  $n$  je dĺžka časového radu a platí  $0 < k \leq n$ .

V časovom rade  $k$ -tu frekvenciu vypočítame a označujeme

$$f_k = \frac{1}{T_k} = \frac{k}{n},$$

Fourierovu frekvenciu, resp.  $k$ -tu uhlovú frekvenciu

$$\omega_k = 2\pi f_k = 2\pi \frac{k}{n}.$$

Najväčšia dĺžka periódy, ktorú sme schopní určiť je rovná počtu pozorovaní t.j.  $T_{max} = n$ , odkiaľ  $k = 1$  a minimálna frekvencia má podľa predpisu veľkosť  $\omega_{min} = 2\pi/n$ .

Najkratšia zistiteľná perióda má dĺžku 2,  $T_{min} = 2$ . Tejto dĺžke potom odpovedá frekvencia  $\omega_{max} = \pi$ . Z predchádzajúcej úvahy vyplýva, že  $k$  môže nadobúdať hodnoty:

$$k = \begin{cases} 1, 2, \dots, n/2 & \text{pre } n \text{ párne} \\ 1, 2, \dots, (n-1)/2 & \text{pre } n \text{ nepárne} \end{cases}$$

Model časovej rady potom môžeme zapísať

$$y'_t = P_t + \epsilon_t,$$

kde

- $t$  ekvidištantné časové okamžiky meraní  $t = t_1, t_2, \dots, t_n$  (pre jednoduchosť predpokladáme, že intervaly majú jednotkovú dĺžku  $t = 1, 2, \dots, n$ )
- $n$  je počet nameraných hodnôt, dĺžka časového radu
- $\epsilon_t$  je biely šum  $\epsilon_t \sim WN(0, \sigma^2)$ ,
- $P_t$  je periodická funkcia tvaru

$$P_t = \begin{cases} \sum_{j=1}^p (\alpha_j \cos t\omega_j + \beta_j \sin t\omega_j) & \text{pre } n \text{ nepárne} \\ \sum_{j=1}^p (\alpha_j \cos t\omega_j + \beta_j \sin t\omega_j) + \alpha_{\frac{n}{2}} (-1)^t & \text{pre } n \text{ párne} \end{cases}$$

kde  $\alpha_j, \beta_j \in \mathbb{R}^1$  odhadujeme metódou najmenších štvorcov a  $p \in N$  je počet periód, ktoré Fisherov test označí ako významné.

Predpis pre  $P_t$  zámerne rozdeľujeme na dva prípady, a to pre  $n$  párne a pre  $n$  nepárne. Treba si uvedomiť, že pre  $n$  párne môže byť medzi periódami aj perióda dĺžky  $T_{min} = 2$  odpovedajúca frekvencii  $\omega_{max} = \pi$ . Nakoľko platí

$$\sin(t\pi) = 0 \quad \cos(t\pi) = (-1)^t,$$

zapisujeme koeficient prislúchajúci tejto frekvencii oddelene.

Periód, resp. frekvencie ktoré zahrnieme do modelu získame pomocou periodogramu a Fisherovho testu, kde vyberieme iba tie periód, ktoré sú významné z hľadiska nášho časového radu. [1][6][9]

### 6.3 Kritéria posúdenia vhodnosti modelu

Po odhadnutí všetkých zložiek a zostavení modelu nás zaujíma nakoľko je daný model dobrý, ako dobre kopíruje dáta. Na to sa používajú rôzne štatistiky.

- **Index determinácie**

Môže nadobúdať hodnoty z intervalu 0 až 1. Čím je hodnota indexu bližšia číslu 1, tým väčšia časť celkovej variability je modelom vysvetlená a tým je model lepší. Vzorec pre výpočet je tvaru

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



- **Reziduálny súčet štvorcov - RSS**

Pre RSS platí, že čím je jeho hodnota menšia, tým model lepšie kopíruje dáta, tým sú odchýlky od skutočnosti menšie.

$$RSS = \sum_{t=1}^n (Y_t - \hat{y}_t)^2 = \sum_{t=1}^n \epsilon_t^2$$

[4][10]

## 7 Rozdelenia pravdepodobností

### 7.1 Normálne rozdelenie

**Definícia 7.1.** Náhodná veličina  $Y$  má normálne rozdelenie s parametrami  $\mu \in \mathbb{R}$  a  $\sigma^2 > 0$ , ak je jej hustota rovná

$$f(Y) = \frac{1}{\sqrt{2\pi}\sigma} \exp \left\{ -\frac{(Y - \mu)^2}{2\sigma^2} \right\}.$$

Zapíšeme  $Y \sim N(\mu, \sigma^2)$  a platí  $EY = \mu$  a  $\text{var}Y = \sigma^2$ .

**Veta 7.1.** Majme  $a, b$  - reálne čísla. Ak  $Y \sim N(\mu, \sigma^2)$ , potom  $a + bY \sim N(a + b\mu, b^2\sigma^2)$ .

**Dôkaz 7.1.** vid' [3] str. 63.

**Definícia 7.2.** Majme náhodný vektor  $\mathbf{Y} = (Y_1, Y_2, \dots, Y_n)^T$  a daný vektor  $\boldsymbol{\mu} = (\mu_1, \mu_2, \dots, \mu_n)^T$ , symetrickú pozitívne semidefinitnú maticu  $\boldsymbol{\Sigma} = (\sigma_{ij})$  typu  $n \times n$ . Vektor  $\mathbf{Y}$  má  $n$ -rozmerné normálne rozdelenie s parametrami  $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$  ak pre ľubovoľný vektor  $\mathbf{c} \in \mathbb{R}^n$  platí  $\mathbf{c}^T \mathbf{Y} \sim N(\mathbf{c}^T \boldsymbol{\mu}, \mathbf{c}^T \boldsymbol{\Sigma} \mathbf{c})$ .

Ak má  $\mathbf{Y}$   $n$ -rozmerné normálne rozdelenie s parametrami  $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ , značíme to  $\mathbf{Y} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ , alebo  $\mathbf{Y} \sim N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

**Veta 7.2.** Ak  $\mathbf{Y} = (Y_1, Y_2, \dots, Y_n)^T \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ , potom  $E\mathbf{Y} = \boldsymbol{\mu}$  a  $\text{var}\mathbf{Y} = \boldsymbol{\Sigma}$ .

**Dôkaz 7.2.** vid' [3] str. 64.

**Veta 7.3.** Nech  $\mathbf{Y} = (Y_1, Y_2, \dots, Y_n)^T \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ . Potom  $\mathbf{X} = \mathbf{a}_{m \times 1} + \mathbf{B}_{m \times n} \mathbf{Y} \sim N(\mathbf{a} + \mathbf{B}\boldsymbol{\mu}, \mathbf{B}\boldsymbol{\Sigma}\mathbf{B}^T)$ .

**Dôkaz 7.3.** vid' [3] str. 64.

**Veta 7.4.** Nech  $\mathbf{Y} \sim N_n(\mathbf{0}, \boldsymbol{\Sigma})$  a nech matica  $\mathbf{A}$  typu  $n \times n$  je symetrická a pozitívne semidefinitná. Ak je matica  $\mathbf{A}\boldsymbol{\Sigma}$  nenulová a idempotentná, tak náhodná veličina  $\mathbf{Y}^T \mathbf{A} \mathbf{Y}$  má  $\chi^2$ -rozdelenie s počtom stupňov voľnosti rovným stope matice  $\mathbf{A}\boldsymbol{\Sigma}$ .

**Dôkaz 7.4.** vid' [3] str. 69.

**Veta 7.5.** Nech  $\mathbf{Y} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$  a nech  $\mathbf{A}$  je matica typu  $n \times n$ , pričom platí  $\mathbf{A} \geq 0$ . Nech  $\mathbf{B}$  je taká matica typu  $m \times n$ , že  $\mathbf{B}\boldsymbol{\Sigma}\mathbf{A} = \mathbf{0}$ . Potom pre ľubovoľný vektor  $\mathbf{c} \in \mathbb{R}^n$  platí, že náhodná veličina  $(\mathbf{Y} - \mathbf{c})^T \mathbf{A}(\mathbf{Y} - \mathbf{c})$  a náhodný vektor  $\mathbf{B}\mathbf{Y}$  sú nezávislé.

**Dôkaz 7.5.** vid' [3] str. 69.

**Veta 7.6.** Nech  $Y_1, Y_2, \dots, Y_n$  je výber z  $N(\mu, \sigma^2)$ . Potom platí:

1.  $\bar{Y} \sim N(\mu, \sigma^2/n)$

2. ak je  $n > 1$  a zároveň  $\sigma^2 > 0$ , potom  $\frac{(n-1)S^2}{\sigma^2} \sim \chi_{n-1}^2$ , kde

$$S^2 = \sum_{i=1}^n (Y_i - \bar{Y})^2 / (n-1)$$

3. pre  $n > 1$  sú veličiny  $\bar{Y}$  a  $S^2$  nezávislé.

**Dôkaz 7.6.** 1. Ak označíme  $\boldsymbol{\mu} = (\mu, \mu, \dots, \mu)^T$ , potom  $\mathbf{Y} = (Y_1, Y_2, \dots, Y_n)^T \sim N(\boldsymbol{\mu}, \sigma^2 \mathbf{I})$ . Na dôkaz rovnosti použijeme vetu (7.3), kde  $\mathbf{a}$  položíme rovné  $\mathbf{0}$  a  $\mathbf{B} = (1/n, 1/n, \dots, 1/n)$ . Odtiaľ dostávame  $E\bar{Y} = \mathbf{a} + \mathbf{B}\boldsymbol{\mu} = \mu/n + \mu/n + \dots + \mu/n = \mu$  a  $\text{var}\bar{Y} = \mathbf{B}\boldsymbol{\Sigma}\mathbf{B}^T = \sigma^2/n^2 + \sigma^2/n^2 + \dots + \sigma^2/n^2 = \sigma^2/n$ .

2. Označme  $\mathbf{Y} = (Y_1, Y_2, \dots, Y_n)$ , kde  $Y_i = (Y_i - \mu)/\sigma$  pre  $i = 1, 2, \dots, n$  a

$$\bar{Y} = \sum_{i=1}^n Y_i / n = (\bar{Y} - \mu) / \sigma. \text{ Dostávame } \mathbf{Y} \sim N(\mathbf{0}, \mathbf{I}) \text{ a } \frac{(n-1)S^2}{\sigma^2} =$$

$$\sigma^{-2} \sum_{i=1}^n (Y_i - \bar{Y})^2 = \sigma^{-2} \sum_{i=1}^n [(Y_i - \mu) - (\bar{Y} - \mu)]^2 = \sum_{i=1}^n (Y_i - \bar{Y})^2.$$

Predchádzajúcu sumu môžeme ešte ďalej rozpísať ako:

$$\sum_{i=1}^n (Y_i - \bar{Y})^2 = \sum_{i=1}^n \left( Y_i - \frac{1}{n} \sum_{j=1}^n Y_j \right)^2 =$$

$$\begin{aligned}
&= \sum_{i=1}^n \left( \frac{1}{n^2} \sum_{j=1}^n \sum_{k=1}^n Y_j Y_k - \frac{2}{n} Y_i \sum_{j=1}^n Y_j + Y_i^2 \right) = \sum_{i=1}^n Y_i^2 - \frac{2}{n} \sum_{i=1}^n \sum_{j=1}^n Y_i Y_j + \\
&+ \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n Y_i Y_j = \sum_{i=1}^n Y_i^2 - \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n Y_i Y_j.
\end{aligned}$$

Označme  $\mathbf{J}$  maticu typu  $n \times n$  takú, že všetky jej prvky sú rovné jednej.

Ďalej označme  $\mathbf{A} = \mathbf{I} - \frac{1}{n}\mathbf{J}$ . Potom môžeme písať

$$(n-1)S^2/\sigma^2 = \sum_{i=1}^n (Y_i - \bar{Y})^2 = \mathbf{Y}^T \mathbf{J} \mathbf{Y} \quad (5)$$

Ľahko sa overí, že matica  $\mathbf{A}$  je symetrická a idempotentná a jej hodnosť je rovná jej stope, teda  $n-1$ . Pretože  $\mathbf{Y} \sim N(0, 1)$ , z vety 7.4 vyplýva  $\mathbf{Y}^T \mathbf{A} \mathbf{Y} \sim \chi_{n-1}^2$ .

3. Po vyjadrení  $S^2$  zo vzorca 5 dostávame

$$S^2 = \frac{\sigma^2}{n-1} \mathbf{Y}^T \mathbf{A} \mathbf{Y} = \frac{1}{n-1} (\mathbf{Y} - \boldsymbol{\mu})^T \mathbf{A} (\mathbf{Y} - \boldsymbol{\mu})$$

Prvý a posledný člen rovnosti sú si rovné aj v prípade  $\sigma^2 = 0$ . Označme znova  $\boldsymbol{\mu} = (\mu, \mu, \dots, \mu)$  a  $\mathbf{B} = (1/n, 1/n, \dots, 1/n)$ . Nakoľko  $\mathbf{Y} \sim N(\boldsymbol{\mu}, \sigma^2 \mathbf{I})$  a  $\bar{Y} = \mathbf{B} \mathbf{Y}$ , môžeme použiť vetu 7.5. Po dosadení nakoniec dostávame

$$\mathbf{B}(\sigma^2 \mathbf{I}) \left( \frac{1}{n-1} \right) \mathbf{A} = 0$$

čím je nezávislosť  $\bar{Y}$  a  $S^2$  dokázaná.

□[3]

## 7.2 Studentovo rozdelenie

**Veta 7.7.** Nech  $X$  a  $Y$  sú nezávislé náhodné veličiny, pričom  $X \sim N(0, 1)$  a  $Y \sim \chi_n^2$ . Náhodná veličina

$$T = \frac{X}{\sqrt{Y/n}} \quad (6)$$

má potom Studentovo  $t$ -rozdelenie s  $n$  stupňami voľnosti (značíme  $t_n$ ).

**Poznámka 7.1.** Kritická hodnota  $t_n(\alpha)$  Studentovho  $t$ -rozdelenia je definovaná vzťahom  $P\{|T| \geq t_n(\alpha)\} = \alpha$ , kde  $T \sim t_n$ .

[3]

### 7.3 F-rozdelenie

**Veta 7.8.** Nech  $X \sim \chi_n^2$  a  $Y \sim \chi_m^2$  a  $X$  a  $Y$  sú nezávislé. Potom

$$Z = \frac{X/n}{Y/m} \sim F_{n,m}.$$

$F_{n,m}$  označuje  $F$ -rozdelenie s  $n$  a  $m$  stupňami voľnosti.

**Poznámka 7.2.** Kritická hodnota  $F_{n,m}(\alpha)$  rozdelenia  $F_{n,m}$  je definovaná vzťahom  $P\{Z \geq F_{n,m}(\alpha)\} = \alpha$ , kde  $Z \sim F_{n,m}$ .

[3]

### 7.4 Rozdelenie kvadratických foriem

**Veta 7.9.** Nech  $X_1, X_2, \dots, X_n$  sú nezávislé náhodné veličiny, každá s  $N(0, 1)$  rozdelením. Potom náhodná veličina  $Y = X_1^2 + X_2^2 + \dots + X_n^2$  má  $\chi_n^2$  rozdelenie.

**Veta 7.10.** Nech  $Z$  a  $W$  sú nezávislé náhodné veličiny, pričom  $Z \sim \chi_r^2$  a  $W \sim \chi_s^2$ . Potom pre súčet  $Z + W$  platí  $Z + W \sim \chi_{r+s}^2$ .

[3]

## 8 Intervaly spoľahlivosti a testovanie hypotéz

### 8.1 Intervaly spoľahlivosti vektoru regresných parametrov

**Veta 8.1.** *Majme lineárny regresný model  $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ , pričom  $\mathbf{Y} \sim N_n(\boldsymbol{\mu}, \sigma^2\mathbf{I})$ ,  $\boldsymbol{\beta}$  je rozmeru  $m \times 1$ . Náhodná veličina  $(n-m)\hat{\sigma}^2/\sigma^2$  má  $\chi^2$  - rozdelenie s  $(n-m)$  stupňami voľnosti a náhodná veličina  $(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^T \mathbf{X}^T \mathbf{X} (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})/\sigma^2$  má  $\chi^2$  - rozdelenie s  $m$  stupňami voľnosti. Podiel týchto nezávislých náhodných veličín, pričom každá je podelená stupňami voľnosti, má  $F$  - rozdelenie s  $m$  a  $(n-m)$  stupňami voľnosti. Pre hranice  $100(1-\alpha)\%$  intervalu spoľahlivosti potom platí*

$$(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^T \mathbf{X}^T \mathbf{X} (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = m\hat{\sigma}^2 F_{1-\alpha}(m, n-m), \quad (7)$$

kde  $F_{1-\alpha}(m, n-m)$  je  $1-\alpha$  kvantil s  $m$  a  $(n-m)$  stupňami voľnosti.

Matica  $\mathbf{X}^T \mathbf{X}$  je regulárna, preto predchádzajúci vzťah definuje hyperelipsoid, ktorého osi sú orientované v smeroch vlastných vektorov matice  $(\mathbf{X}^T \mathbf{X})^{-1}$ . Jednotlivé poloosi majú dĺžku rovnú  $p\sqrt{\lambda_j}$ , pričom  $\lambda_j$  sú vlastné čísla matice  $(\mathbf{X}^T \mathbf{X})^{-1}$  a  $p = \sqrt{m\hat{\sigma}^2 F_{1-\alpha}(m, n-m)}$ .

Pri odhadovaní vektoru parametrov  $\boldsymbol{\beta}$  musíme pamätať na to, že ide iba o bodové odhady. Tieto majú v praxi omnoho menší význam ako konfidenčné intervaly, nazývané tiež oblasti alebo intervaly spoľahlivosti. V týchto leží skutočná hodnota parametra  $\boldsymbol{\beta}$  s dopredu zvolenou pravdepodobnosťou  $(1-\alpha)$ . Pravdepodobnosť pokrytia sa nazýva hladina spoľahlivosti a obyčajne sa volí 95% alebo 99%.

Častejšie ako zostrojovanie intervalov spoľahlivosti pre celý vektor regresných parametrov zostrojujeme intervaly pre jednotlivé zložky. Tento postup je omnoho prehľadnejší a úzko súvisí s testovaním významnosti regresných parametrov vysvetleným v nasledujúcej podkapitole, nakoľko ak interval spoľahlivosti obsahuje 0, parameter považujeme za nevýznamný a odmietne ho aj test významnosti.

**Veta 8.2.** *Nech je korelácia medzi odhadmi parametrov nulová. Potom  $100(1 - \alpha)\%$  jednoduché intervaly spoľahlivosti pre parameter  $\beta_j$  môžeme odvodiť zo  $100(1 - \alpha)\%$  intervalov pre celý vektor  $\beta$  a budú mať tvar*

$$\left\langle \hat{\beta}_j - t_{1-\alpha/2}(n - m)\hat{\sigma}\sqrt{v_{jj}}, \hat{\beta}_j + t_{1-\alpha/2}(n - m)\hat{\sigma}\sqrt{v_{jj}} \right\rangle, \quad (8)$$

kde  $v_{jj}$  sú diagonálne prvky matice  $(\mathbf{X}^T\mathbf{X})^{-1}$  a  $\hat{\beta}_j$  odhady zložiek parametra  $\beta$ .

[13]

## 8.2 Testovanie významnosti regresných parametrov

Pri testovaní hypotéz rôznych parametrov vychádzame z predpokladu, že nami skúmaný parameter, napr. stredná hodnota  $\mu$  je rovná určitej hodnote, napr. 0. Nevieme však, či je toto tvrdenie pravdivé, preto ho nazveme hypotézou. Pri testovaní sa táto hypotéza zvyčajne nazýva nulová a označuje sa  $H_0$ . Proti hypotéze  $H_0$  stojí alternatívna hypotéza  $H_1$  alebo  $H_a$ , ktorá je akýmsi doplnením nulovej hypotézy. Teda ide o hypotézu ktorá platí práve vtedy, ak neplatí nulová hypotéza.

Chyba, ktorej sa môžeme dopustiť ak nulová hypotéza platí a my ju zamietneme sa nazýva chyba prvého druhu. Pravdepodobnosť tejto chyby je rovná dopredu zvolenému  $\alpha$ , ktoré sa nazýva aj hladina významnosti testu. Testovanie hypotéz úzko súvisí s konštrukciou intervalov spoľahlivosti.

Pre vektor regresných parametrov  $\beta$  testujeme najčastejšie nulovú hypotézu v tvare  $H_0 : \beta = \beta_0$  proti alternatívnej hypotéze  $H_1 : \beta \neq \beta_0$ , kde  $\beta_0$  je väčšinou rovné nulovému vektoru. Testovaciu štatistiku ľahko odvodíme zo vzorca 7 a bude v tvare

$$F = \frac{(\hat{\beta} - \beta_0)^T \mathbf{X}^T \mathbf{X} (\hat{\beta} - \beta_0)}{m\hat{\sigma}^2}$$

a za platnosti nulovej hypotézy bude mať  $F$  – rozdelenie s  $m$  a  $n - m$  stupňami voľnosti, kde  $n$  je dĺžka časového radu a  $m$  je počet odhadovaných parametrov modelu, resp. dimenzia vektoru  $\beta$ .

Nulovú hypotézu ( $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0$ ) zamietame ak pre testovaciu štatistiku  $F$  platí :  
 $|F| \geq F_{m,n-m}(1 - \alpha/2)$ .

Ak by sme testovali jednotlivé zložky  $\hat{\beta}_j$  parametra  $\hat{\boldsymbol{\beta}}$ , pri nulovej hypotéze  $H_0 : \hat{\beta}_j = \beta_{j_0}$ , testová štatistika by mala tvar:

$$T_j = \frac{|\hat{\beta}_j - \beta_{j_0}|}{\hat{\sigma} \sqrt{v_{jj}}},$$

ktorá má za platnosti nulovej hypotézy Studentovo rozdelenie s  $n - m$  stupňami voľnosti,  $v_{jj}$  vo vzorci sú diagonálne prvky matice  $(\mathbf{X}^T \mathbf{X})^{-1}$ . Nulovú hypotézu zamietame ak  $|T_j| \geq t_{n-m}(1 - \alpha/2)$ .

[13] [3]

### 8.3 Jednovýberový t-test

Nech  $X_1, X_2, \dots, X_n$  je náhodný výber z  $N(\mu, \sigma^2)$  rozdelenia, pričom  $n \geq 2$  a  $\sigma^2 > 0$ . Predpokladajme, že parametre  $\mu$  a  $\sigma^2$  sú neznáme. Testujme hypotézu  $H_0 : \mu = \mu_0$  proti alternatíve  $H_1 : \mu \neq \mu_0$ .

**Veta 8.3.** *Ak je stredná hodnota normálneho rozdelenia skutočne rovná  $\mu$  potom náhodná veličina*

$$T = \frac{\bar{X} - \mu}{S} \sqrt{n} \sim t_{n-1}$$

*má  $t_{n-1}$  rozdelenie.*

Nulovú hypotézu  $H_0$  zamietame ak platí  $|\bar{X} - \mu_0| \sqrt{n}/S \geq t_{n-1}(\alpha)$ .

[2]

**Dôkaz 8.1.** *vid' [2] str. 85*



## 9 Predikcie

Skúmanie časových radov väčšinou vedie ku konštrukciám predpovedí a analýze budúceho vývoja. Analytici sa pomocou minulosti a súčasnosti snažia predikovať budúcnosť a tak získavajú možnosť sa na ňu čo najlepšie pripraviť. Či už ide o oblasť investovania alebo úplne inú, napr. výrobu, počasie, atď., stále je lepšie mať aspoň čiastočné informácie, aby sme vedeli, čo približne môžeme očakávať.

V štatistike existujú napríklad bodové predpovede, ktoré však nemusia byť úplne presné a preto sa využívajú skôr predpovede intervalové, ktoré s určitou pravdepodobnosťou pokrývajú skutočnú hodnotu.

### 9.1 Bodové predpovede

**Definícia 9.1.** *Nech  $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$  je lineárny regresný model. Bodová predpoveď  $\hat{y}_0$  v nejakom budúcom bode  $\mathbf{t}_0 = (t_{01}, t_{02}, \dots, t_{0k})^T$  je vyjadrená vzťahom  $\hat{y}_0 = \mathbf{t}_0^T \hat{\boldsymbol{\beta}}$ .*

Bodové predpovede konštruujeme väčšinou iba pre prvých niekoľko nasledujúcich časových okamžikov časového radu.

[14]

### 9.2 Intervalové predpovede

Konštrukcia intervalových predpovedí, resp. intervalov spoľahlivosti pre predpovede, predikčných intervalov je podobná konštrukcii intervalov spoľahlivosti pre parameter  $\boldsymbol{\beta}$ .

**Veta 9.1.** *Nech  $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$  je lineárny regresný model a  $\hat{y}_0 = \mathbf{t}_0^T \hat{\boldsymbol{\beta}}$  je bodová predpoveď zostrojená v bode  $\mathbf{t}_0$ . Interval spoľahlivosti s hladinou spoľahlivosti  $1 - \alpha$  pre  $y_0$  má tvar*

$$\left\langle \mathbf{t}_0^T \hat{\boldsymbol{\beta}} - t_{n-k}(1 - \alpha/2)\hat{\sigma}\sqrt{1 + \mathbf{t}_0^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{t}_0}, \mathbf{t}_0^T \hat{\boldsymbol{\beta}} + t_{n-k}(1 - \alpha/2)\hat{\sigma}\sqrt{1 + \mathbf{t}_0^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{t}_0} \right\rangle.$$

**Dôkaz 9.1.** Podľa vety (7.7) vzniká náhodná veličina so studentovým rozdelením ako podiel nezávislých náhodných veličín s  $N(0, 1)$  rozdelením a odmocniny z náhodnej veličiny s  $\chi^2$  - rozdelením podelenej stupňami voľnosti.

Náhodná veličina  $y_0 - \hat{y}_0$  má  $N(0, \sigma^2(1 + \mathbf{t}_0^T(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{t}_0))$  - rozdelenie, pretože platí:

$$E(y_0 - \hat{y}_0) = E(y_0) - E(\hat{y}_0) = 0$$

$$\text{var}(y_0 - \hat{y}_0) = \text{var}(y_0) + \text{var}(\hat{y}_0) - 2\text{cov}(y_0, \hat{y}_0) = \sigma^2 + \sigma^2\mathbf{t}_0^T \text{var}(\hat{\boldsymbol{\beta}})\mathbf{t}_0 = \sigma^2(1 + \mathbf{t}_0^T(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{t}_0)$$

a  $\hat{y}_0$  vzniká ako predikcia normálne rozdelených veličín.

Aby sme získali  $N(0, 1)$  - rozdelenie musíme danú náhodnú veličinu znormovať:

$$\frac{y_0 - \hat{y}_0 - E(y_0 - \hat{y}_0)}{\sqrt{\text{var}(y_0 - \hat{y}_0)}} = \frac{\hat{y}_0 - \mathbf{t}_0^T \hat{\boldsymbol{\beta}}}{\sqrt{\sigma^2(1 + \mathbf{t}_0^T(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{t}_0)}} \sim N(0, 1)$$

Veličina  $\hat{\sigma}^2 = \frac{(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})^T(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})}{n - k}$  má  $\frac{\sigma^2}{n - k}\chi_{n-k}^2$  rozdelenie, odkiaľ

$$\hat{\sigma}^2(n - k)/\sigma^2 \sim \chi_{n-k}^2.$$

Dosadením týchto dvoch náhodných veličín do vzorca 6 získavame náhodnú veličinu so Studentovým  $t$ -rozdelením s  $n - k$  stupňami voľnosti.

$$\frac{\hat{y}_0 - \mathbf{t}_0^T \hat{\boldsymbol{\beta}} / \sqrt{\sigma^2(1 + \mathbf{t}_0^T(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{t}_0)}}{\sqrt{\hat{\sigma}^2(n - k) / \sigma^2(n - k)}} = \frac{\hat{y}_0 - \mathbf{t}_0^T \hat{\boldsymbol{\beta}}}{\hat{\sigma} \sqrt{(1 + \mathbf{t}_0^T(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{t}_0)}} \sim t_{n-k}.$$

Odtiaľ platí

$$P \left[ -t_{n-k}(1 - \alpha/2) \leq \frac{\hat{y}_0 - \mathbf{t}_0^T \hat{\boldsymbol{\beta}}}{\hat{\sigma} \sqrt{(1 + \mathbf{t}_0^T(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{t}_0)}} \leq t_{n-k}(1 - \alpha/2) \right] = 1 - \alpha.$$

Riešením danej nerovnosti dostávame hľadané predikčné intervaly.

[14]

## Praktická časť

### 10 Štatistická analýza

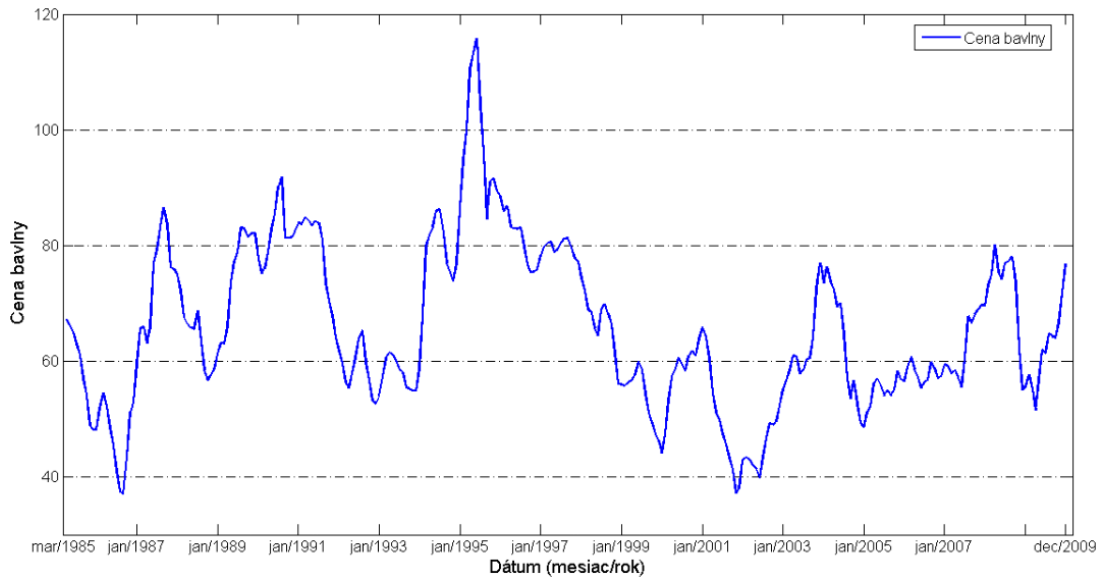
#### 10.1 Časový rad ceny bavlny na burze

V praktickej časti sa budeme venovať konkrétnym komoditám a vývoju ich ceny na burze za posledných približne 25 rokov. V predchádzajúcich kapitolách sme si uviedli teoretické východiská, ktoré teraz aplikujeme na konkrétne dáta. Vývoj cien komodít na burze je pri prvom pohľade na cenovú krivku prislúchajúcu určitej komodite veľmi chaotický a nepravidelný. Napriek tomuto faktoru sa budeme snažiť analyzovať vývoj ceny bavlny od marca 1985 do decembra 2009. K dispozícii máme mesačné dáta, teda celkovo 298 dát. Naším cieľom je zistiť trend vývoja ceny, či daný časový rad obsahuje skryté periodicity a ak áno čo najvernejšie popísať priebeh pomocou modelu skrytých períód a predikovať budúci vývoj. Taktiež určiť konfidenčné intervaly, v ktorých sa naša predikcia bude pohybovať a porovnať náš odhad s realitou. Daný postup potom vyskúšame aj na časovom rade vývoja ceny kakaových bôbov.

Cena bavlny na burze sa udáva v amerických centoch na libru (1 libra = 0,4536 kg). Zdrojom dát je webová stránka <http://www.indexmundi.com> [26]. Všetky použité dáta sú uvedené v prílohe E.

Na obrázku 1 je graficky znázornený vývoj ceny bavlny od marca 1985 do decembra 2009. Po prvom pohľade na dáta je ťažké s istotou určiť trend vývoja. V kapitole 6.1 sme si uviedli rôzne druhy trendových funkcií. Napríklad konštantnú, lineárnu, kvadratickú, ale aj druhý, vo väčšine prípadov výhodnejší postup hľadania trendovej funkcie - polynomicnú regresiu pomocou Čebyševovej metódy. Rozhodli sme sa preto tento postup vyskúšať aj v našom prípade.

Pomocou vzorcov uvedených v kapitole 6.1.4 sme si vypočítali hodnoty polynómov  $\phi_0(t)$  až  $\phi_4(t)$  pre každý časový okamžik  $t = 1, 2, \dots, 298$ . Zostavili sme päť trendových modelov, pričom sme zakaždým zvýšili stupeň polynómu o jeden.



Obr. 1: Vývoj ceny bavlny od marca 1985 do decembra 2009

Prvý model má tvar

$$T_t = \phi_0(t)\delta_0, \quad t = 1, 2, \dots, 298,$$

až posledný je tvaru

$$T_t = \phi_0(t)\delta_0 + \phi_1(t)\delta_1 + \phi_2(t)\delta_2 + \phi_3(t)\delta_3 + \phi_4(t)\delta_4, \quad t = 1, 2, \dots, 298.$$

Pre každý model sme si vypočítali pomocou metódy najmenších štvorcov uvedenej v kapitole 5.1 odhady parametrov  $\delta_i$ . Postup si ukážeme na poslednom modeli zahrňujúcom polynómy  $\phi_0$  až  $\phi_4$ . Môžeme ho maticovo zapísať ako  $\mathbf{T}_t = \mathbf{X}\boldsymbol{\delta}$ , kde stĺpce matice  $\mathbf{X}$  sú postupne vektory  $\phi_0(t)$  až  $\phi_4(t)$  a vektor  $\boldsymbol{\delta} = (\delta_0, \delta_1, \delta_2, \delta_3, \delta_4)^T$ . Odhad vektora  $\boldsymbol{\delta}$  označíme  $\hat{\boldsymbol{\delta}}$  a vypočítame zo vzťahu

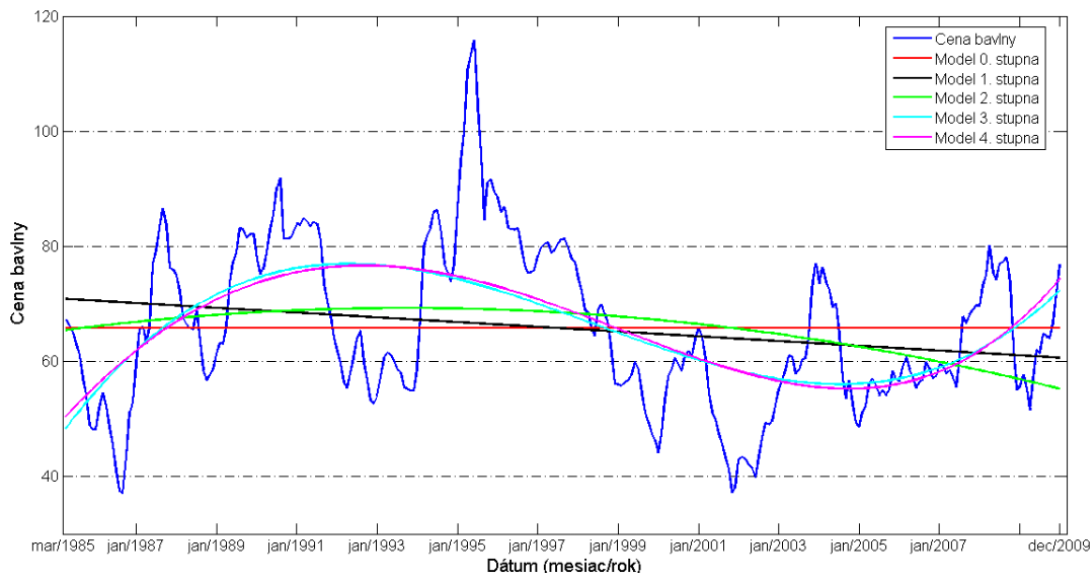
$$\hat{\boldsymbol{\delta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}_t,$$

kde  $\mathbf{y}_t$  je vektor pôvodného časového radu.

Po vypočítaní odhadov parametrov  $\boldsymbol{\delta}$ , môžeme odhadnúť trendovú funkciu

$$\hat{\mathbf{T}}_t = \mathbf{X}\hat{\boldsymbol{\delta}}.$$

Tento postup sme aplikovali na všetkých 5 zostavených modelov a odhady trendových funkcií sme zakreslili do nasledujúceho grafu (obr. 2). Jednotlivé grafy osobitne je možné nájsť v prílohe A.



Obr. 2: Trendové funkcie

Graficky najlepšie kopíruje dáta model s polynómami až do 4. stupňa, ale túto svoju domnienku musíme matematicky podporiť. Najprv si všetkých 5 modelov porovnáme na základe indexu determinácie a reziduálneho súčtu štvorcov (tabuľka 1).

Stupeň modelu	Index determinácie - $R^2$	Reziduálny súčet štvorcov
0.	0	57782.22
1.	0.045202	55170.34
2.	0.07603	53389.04
3.	0.30039	40424.87
4.	0.30298	40275.41

Tabuľka 1: Porovnanie modelov 0. až 4. stupňa

Aj podľa indexu determinácie, aj podľa reziduálneho súčtu štvorcov sa stále najlepšie javí model 4. stupňa. Aby sme však tento model naozaj mohli označiť za najlepší odhad našej trendovej funkcie, musíme ešte otestovať významnosť parametrov  $\delta_i$ . To urobíme pre všetkých päť modelov.

Položíme nulovú hypotézu

$$H_0 : \delta_i = 0 \quad i = 0, 1, 2, 3, 4$$

proti alternatíve

$$H_1 : \delta_i \neq 0, \quad i = 0, 1, 2, 3, 4.$$

Testová štatistika je tvaru

$$T = \frac{\delta_i - 0}{\hat{\sigma} \sqrt{v_{ii}}}$$

a ak  $|T| \geq t_{298-m}(1 - \alpha/2)$ , kde  $m$  je počet odhadovaných parametrov modelu, tak hypotézu  $H_0$  zamietame, čiže parameter je významný,  $v_{ii} = \{(\mathbf{X}^T \mathbf{X})^{-1}\}_{ii}$ .

V tabuľke 2 si uvedieme odhady parametra  $\hat{\sigma}^2 = (\mathbf{y}_t - \hat{\mathbf{T}}_t)^T (\mathbf{y}_t - \hat{\mathbf{T}}_t) / (298 - m)$  pre jednotlivé modely, kde  $m$  je počet odhadovaných parametrov, čiže postupne  $m = 1, 2, \dots, 5$ .

Stupeň modelu	$\hat{\sigma}^2$
0.	194.55
1.	186.39
2.	180.98
3.	137.5
4.	137.46

Tabuľka 2: Odhady parametra  $\sigma^2$  pre jednotlivé trenové funkcie

A teraz si už môžeme vypočítať hodnoty jednotlivých testovacích štatistík  $T$  pre jednotlivé modely a parametre  $\delta_i$  (tab. 3).

Hodnoty parametrov		Hodnota testovacej štatistiky T / rozhodnutie o zamietnutí				
		Model 0.stupňa	Model 1.stupňa	Model 2.stupňa	Model 3.stupňa	Model 4.stupňa
$\hat{\delta}_0$	65.706	81.32 /zam.	83.08 /zam.	84.31 /zam.	96.73 /zam.	96.74 /zam.
$\hat{\delta}_1$	-0.034		-3.74 /zam.	-3.80 /zam.	-4.36 /zam.	-4.36 /zam.
$\hat{\delta}_2$	-0.000369			-3.14 /zam.	-3.60 /zam.	-3.60 /zam.
$\hat{\delta}_3$	1.319e-005				9.71 /zam.	9.71 /zam.
$\hat{\delta}_4$	1.886e-008					1.04 /nezam.

Tabuľka 3: Hodnoty testovacej štatistiky T

Z predchádzajúcej tabuľky vyplýva, že síce model 4. stupňa má najmenší RSS a najväčší index determinácie zo všetkých skúmaných modelov, jeho použitie nie je nutné, nakoľko odhad regresného parametra  $\delta_4$  sa testom ukázal ako nevýznamný. Odhadom trenovej funkcie bude teda model 3. stupňa

$$\hat{T}_t = \phi_0(t)\hat{\delta}_0 + \phi_1(t)\hat{\delta}_1 + \phi_2(t)\hat{\delta}_2 + \phi_3(t)\hat{\delta}_3, \quad t = 1, 2, \dots, 298$$

$$\hat{T}_t = 65.706\phi_0(t) - 0.034\phi_1(t) - 0.000369\phi_2(t) + 0.00001319\phi_3(t)$$

graficky znázornený na obr. 3.

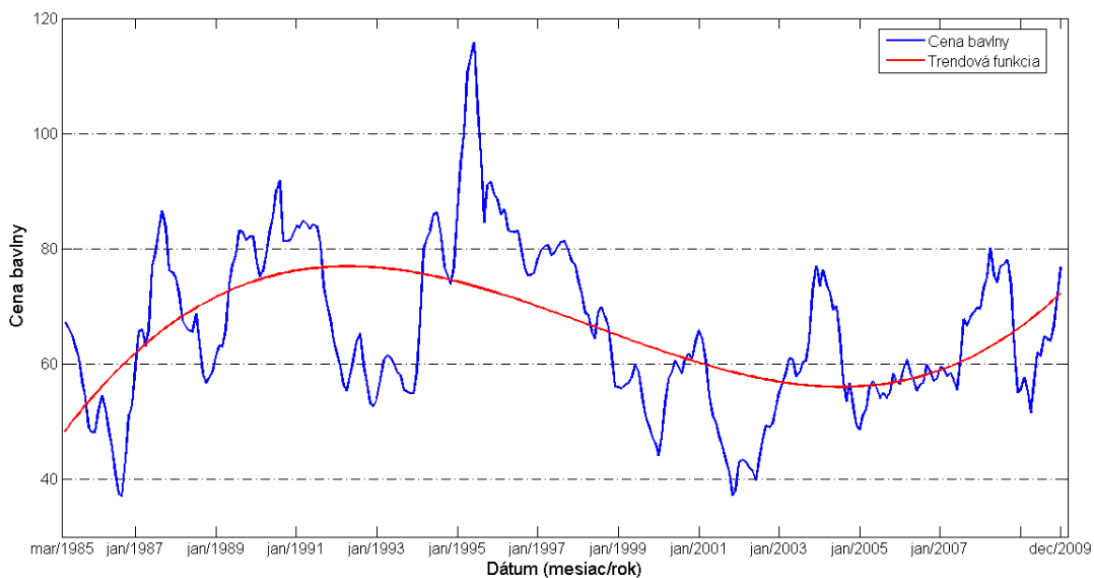
Keďže sme našli odhad trenovej funkcie, môžeme náš časový rad očistiť od trenovej zložky a namiesto časového radu

$$y_t = T_t + S_t + C_t + \epsilon_t$$

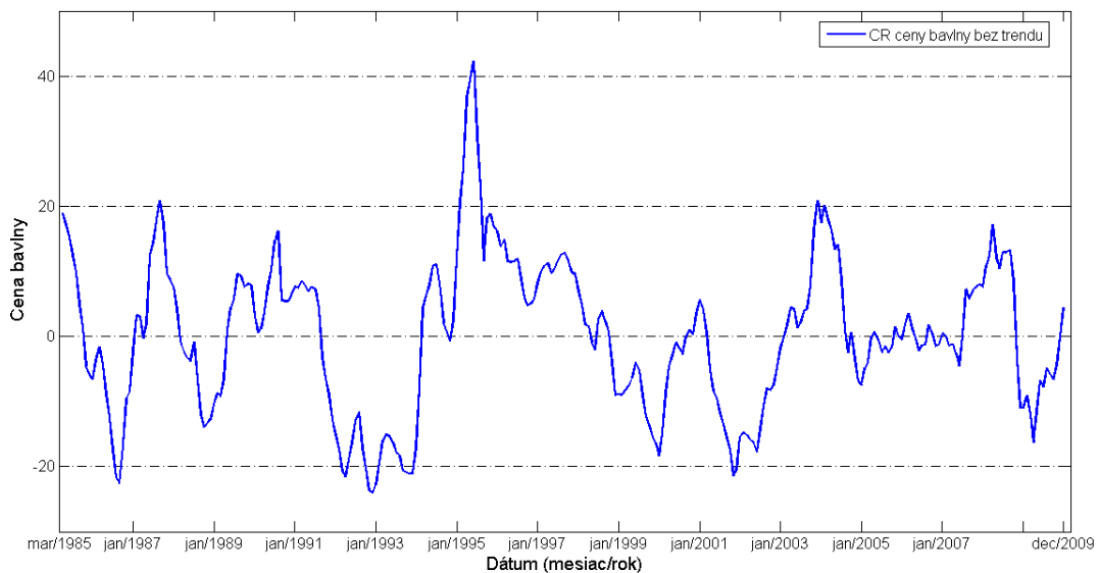
dostávame časový rad tvaru

$$y'_t = y_t - \hat{T}_t = P_t + \epsilon_t, \quad P_t = S_t + C_t$$

graficky na obr. 4.



Obr. 3: Odhad trendovej funkcie



Obr. 4: Časový rad po odpočítaní trendovej zložky

Už na prvý pohľad pozorujeme určitú periodickosť vývoja ceny bavlny v očistenom časovom rade. Je však obtiažne presne odhadnúť dĺžky períód. Preto, aby sme odhalili aké periódy sa v našom časovom rade vyskytujú, použijeme postup nazývaný periodogram bližšie vysvetlený v kapitole 6.2.1.

Na výpočet hodnôt periodogramu sme použili vzorec, ktorý je pre výpočty naj-

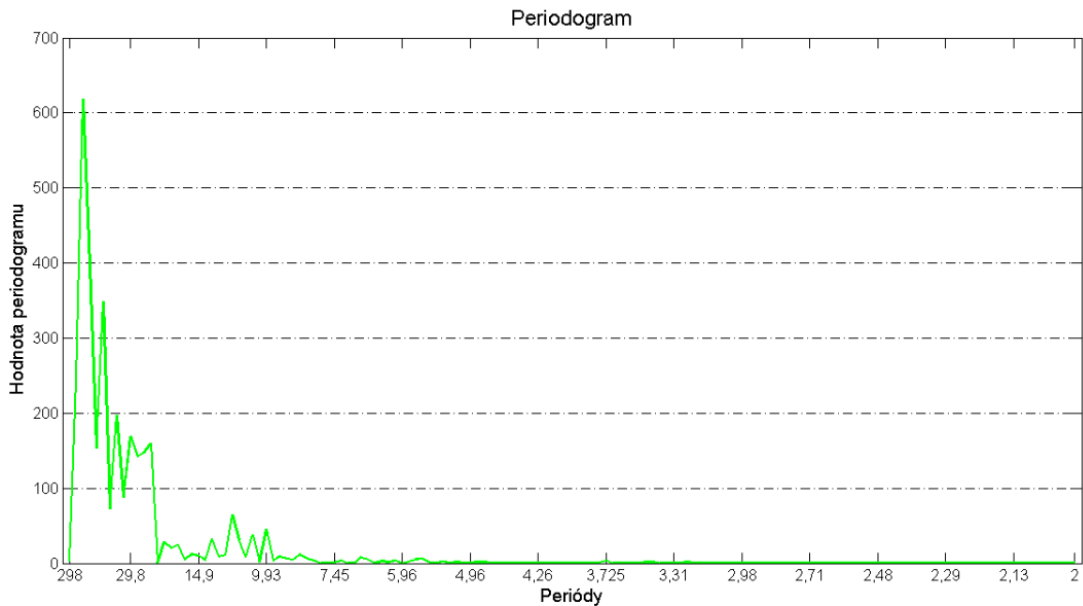


jednoduchší

$$A(\omega) = (2/n)^{1/2} \sum_{t=1}^n Y_t \cos t\omega, \quad B(\omega) = (2/n)^{1/2} \sum_{t=1}^n Y_t \sin t\omega$$

$$I(\omega) = \frac{1}{4\pi} A^2(\omega) + \frac{1}{4\pi} B^2(\omega),$$

kde v našom prípade  $n = 289$ ,  $\omega = 2\pi j/298$ , pričom  $j = 1, 2, \dots, 149$ ,  $Y_t = y_t'$  a  $t = 1, 2, \dots, 298$ . Vypočítané hodnoty periodogramu  $I(\omega)$  prislúchajúce jednotlivým periódam sme zakreslili do grafu (obr.5 ). Všetky číselné hodnoty periodogramu uvádzame v prílohe E.



Obr. 5: Periodogram

Ako vidíme na grafe, najväčšie hodnoty dosahuje periodogram skôr pre veľké periódny ako pre menšie. Hodnoty periodogramu si teraz otestujeme Fisherovým testom, aby sme zistili, ktoré periódny sú z hľadiska nášho časového radu významné. Najprv si musíme všetky hodnoty periodogramu  $I(\omega)$  zoradiť zostupne. Ako prvú budeme testovať najväčšiu hodnotu. Tá prislúcha dĺžke periódny 298/3 a má hodnotu 618.04 . Vypočítame si hodnotu Fisherovej štatistiky

$$W = \frac{V_1}{V_1 + V_2 + \dots + V_{149}} = 0.19209$$

A kritickú hodnotu  $x$  aproximovanú podľa vzorca

$$P(W > x) \cong m(1 - x)^{m-1} = 149(1 - x)^{148} = 0.01.$$

Schválne sme zvolili kritickú hodnotu 0.01 nakoľko vieme, že tento test nie je príliš silný.

Pre prvú najvyššiu hodnotu periodogramu je kritická hodnota  $x = 0.062864$ . Vidíme, že štatistika  $W > x$ , preto hypotézu

$$H_0 : Y_t = \epsilon_t, \quad \epsilon_t \sim WN(0, \sigma^2)$$

zamietame a periódu dĺžky 298/3 považujeme za významnú.

Rovnaký postup zopakujeme pre druhú najvyššiu hodnotu, pričom prvú hodnotu vynechávame. Fisherovu štatistiku vypočítame zo vzťahu

$$W = \frac{V_2}{V_2 + V_3 + \dots + V_{149}} = 0.15258$$

a kritickú hodnotu

$$P(W > x) \cong m(1 - x)^{m-1} = 148(1 - x)^{147} = 0.01,$$

odkiaľ  $x = 0.063235$ . Nulovú hypotézu opäť zamietame a periódu prehlasujeme za významnú. V tabuľke 4 uvádzame hodnoty Fisherovej štatistiky a kritickej hodnoty postupne pre všetky (zostupne radené) hodnoty periodogramu až po prvú, ktorú test nezamietne. Tá je zvýraznená.

	Dĺžka periódy	Fisherová štatistika	Kritická hodnota		Dĺžka periódy	Fisherová štatistika	Kritická hodnota
1	99.33	0.19209	0.062864	12	42.57	0.13069	0.067214
2	74.5	0.15258	0.063235	13	11.92	0.13498	0.067641
3	49.67	0.15844	0.06361	14	9.93	0.10947	0.068075
4	149	0.13192	0.06399	15	10.64	0.10406	0.068513
5	37.25	0.12272	0.064375	16	13.55	0.0976	0.068958
6	29.8	0.12002	0.064765	17	11.46	0.10519	0.069409
7	22.92	0.12864	0.06516	18	19.87	0.10581	0.069867
8	59.6	0.14121	0.065561	19	17.53	0.1025	0.07033
9	24.83	0.15827	0.065966	20	18.63	0.09413	0.070801
10	27.09	0.18197	0.066377	21	15.68	0.061452	0.071277
11	33.11	0.13839	0.066793				

Tabuľka 4: Testovanie významnosti periód

Z uvedeného vyplýva, že do modelu skrytých periód by sme mali zahrnúť prvých 20 periód - periódy prislúchajúce prvým 20 najväčším hodnotám periodogramu. Periodická zložka charakterizovaná modelom skrytých periód bude vyzerat'

$$P_t = \sum_{i=1}^{20} \left( \beta_{2i-1} \cos \frac{2\pi t}{\tau_i} + \beta_{2i} \sin \frac{2\pi t}{\tau_i} \right) + \epsilon_t$$

kde

$\tau_1$	99.33	$\tau_2$	74.5	$\tau_3$	49.67	$\tau_4$	149
$\tau_5$	37.25	$\tau_6$	29.8	$\tau_7$	22.92	$\tau_8$	59.6
$\tau_9$	24.83	$\tau_{10}$	27.09	$\tau_{11}$	33.11	$\tau_{12}$	42.57
$\tau_{13}$	11.92	$\tau_{14}$	9.93	$\tau_{15}$	10.64	$\tau_{16}$	13.55
$\tau_{17}$	11.46	$\tau_{18}$	19.87	$\tau_{19}$	17.53	$\tau_{20}$	18.63

Tabuľka 5: Hodnoty parametrov  $\tau_i$

Odhad parametra  $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_{20})^T$  v tomto modeli získame metódou najmenších štvorcov.

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}_1^T \mathbf{X}_1)^{-1} \mathbf{X}_1^T \mathbf{y}_t'$$

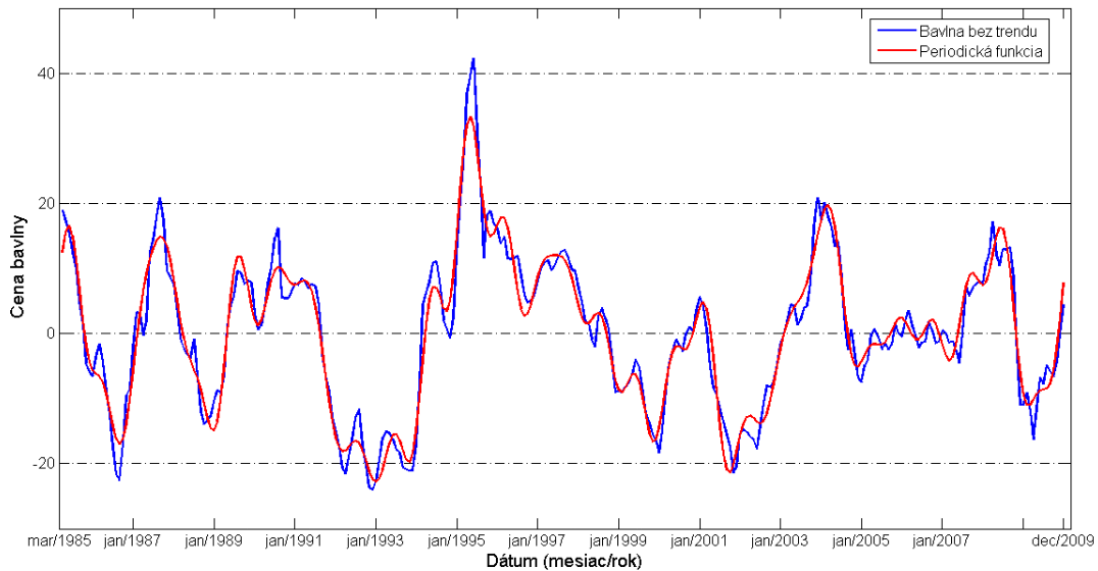
kde matica

$$\mathbf{X}_1 = \begin{bmatrix} \cos \frac{2\pi 1}{99.33} & \sin \frac{2\pi 1}{99.33} & \dots & \cos \frac{2\pi 1}{18.63} & \sin \frac{2\pi 1}{18.63} \\ \cos \frac{2\pi 2}{99.33} & \sin \frac{2\pi 2}{99.33} & \dots & \cos \frac{2\pi 2}{18.63} & \sin \frac{2\pi 2}{18.63} \\ \cdot & \cdot & \dots & \cdot & \cdot \\ \cdot & \cdot & \dots & \cdot & \cdot \\ \cos \frac{2\pi 298}{99.33} & \sin \frac{2\pi 298}{99.33} & \dots & \cos \frac{2\pi 298}{18.63} & \sin \frac{2\pi 298}{18.63} \end{bmatrix}$$

a vektor  $\mathbf{y}_t'$  je vektor hodnôt časového radu ceny bavlny očistený od trendovej zložky.

Pre odhad funkcie periodickej zložky potom dostávame

$$\hat{\mathbf{P}}_t = \mathbf{X}_1 \hat{\boldsymbol{\beta}}.$$



Obr. 6: Odhad periodickej zložky

Na obrázku 6 vidíme odhad periodickej zložky modelu aj to, že veľmi dobre kopíruje dáta. Musíme sa však pozrieť aj na to, či náhodou niektoré z parametrov  $\beta_1$  až  $\beta_{40}$  nie sú nevýznamné. Postup testu bude rovnaký ako pri testovaní významnosti parametrov trendovej funkcie. Testujeme nulovú hypotézu

$$H_0 : \beta_i = 0 \quad \text{proti} \quad H_1 : \beta_i \neq 0$$

Testová štatistika je tvaru

$$T = \frac{\beta_i - 0}{\hat{\sigma}\sqrt{c_{ii}}},$$

kde  $c_{ii}$  sú diagonálne prvky matice  $(\mathbf{X}_1^T \mathbf{X}_1)^{-1}$  a  $\hat{\sigma}^2 = (y_t' - \hat{P}_t)^T (y_t' - \hat{P}_t) / (258) = 9.4366$ . Testovú štatistiku porovnáваме s kritickou hodnotou  $t_{1-0.05/2}(258) = t_{0.975}(258) = 1.9692$ . Ak  $|T| \leq 1.9692$  hypotézu  $H_0$  nezamietame, a parameter považujeme za nevýznamný a môžeme ho z modelu vylúčiť.

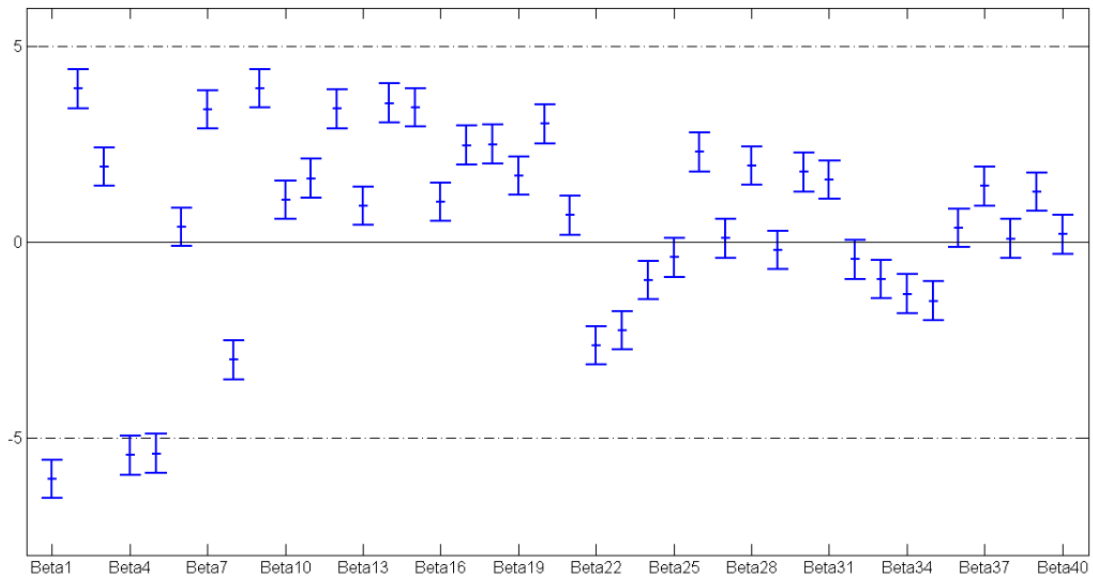
Parameter	T	Rozhodnutie	Parameter	T	Rozhodnutie
$\beta_1$	-24.0763	zamietame	$\beta_{21}$	2.7248	zamietame
$\beta_2$	15.5996	zamietame	$\beta_{22}$	-10.5137	zamietame
$\beta_3$	7.6469	zamietame	$\beta_{23}$	-8.9916	zamietame
$\beta_4$	-21.6724	zamietame	$\beta_{24}$	-3.8909	zamietame
$\beta_5$	-21.5020	zamietame	$\beta_{25}$	-1.5532	nezamietame
$\beta_6$	1.5531	nezamietame	$\beta_{26}$	9.1524	zamietame
$\beta_7$	13.4976	zamietame	$\beta_{27}$	0.3685	nezamietame
$\beta_8$	-11.9775	zamietame	$\beta_{28}$	7.7668	zamietame
$\beta_9$	15.6386	zamietame	$\beta_{29}$	-0.8142	nezamietame
$\beta_{10}$	4.2922	zamietame	$\beta_{30}$	7.1075	zamietame
$\beta_{11}$	6.4878	zamietame	$\beta_{31}$	6.3192	zamietame
$\beta_{12}$	13.5477	zamietame	$\beta_{32}$	-1.7536	nezamietame
$\beta_{13}$	3.6653	zamietame	$\beta_{33}$	-3.7750	zamietame
$\beta_{14}$	14.1201	zamietame	$\beta_{34}$	-5.2513	zamietame
$\beta_{15}$	13.6645	zamietame	$\beta_{35}$	-5.9645	zamietame
$\beta_{16}$	4.1027	zamietame	$\beta_{36}$	1.4393	nezamietame
$\beta_{17}$	9.8401	zamietame	$\beta_{37}$	5.6994	zamietame
$\beta_{18}$	9.9554	zamietame	$\beta_{38}$	0.3600	nezamietame
$\beta_{19}$	6.7257	zamietame	$\beta_{39}$	5.1256	zamietame
$\beta_{20}$	12.0159	zamietame	$\beta_{40}$	0.7791	nezamietame

Tabuľka 6: Testovanie významnosti parametrov modelu skrytých periód

Z tabuľky 6 vidieť, že v modeli máme 8 nepotrebných parametrov, ktoré môžeme vypustiť. Toto tvrdenie nám potvrdili aj intervaly spoľahlivosti (obr. 7) zostrojené podľa vzťahu

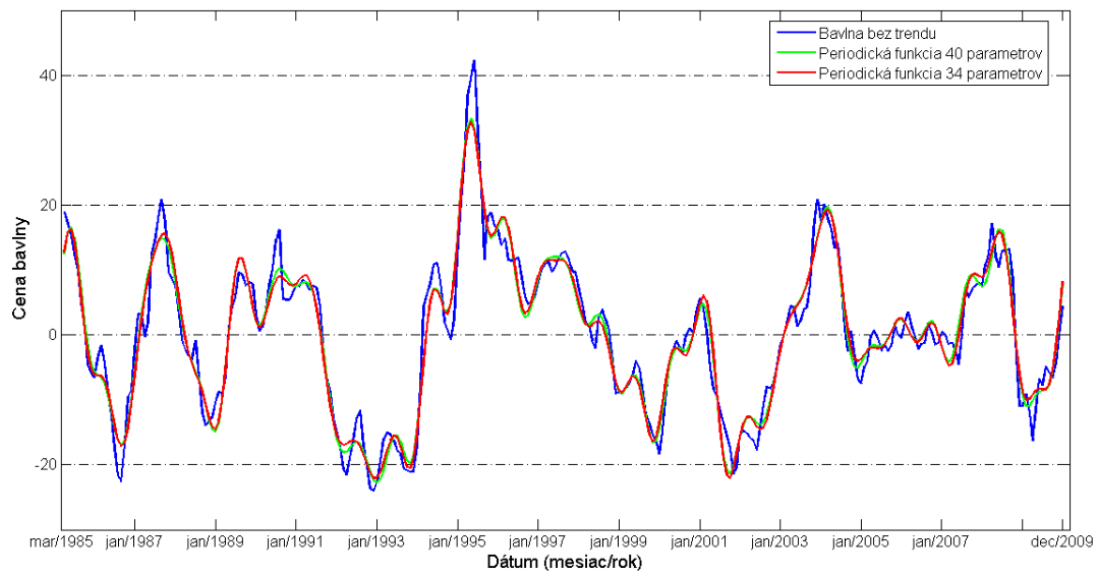
$$\langle \beta_i - t_{1-\alpha/2}(n-m)\hat{\sigma}\sqrt{c_{ii}}, \beta_i + t_{1-\alpha/2}(n-m)\hat{\sigma}\sqrt{c_{ii}} \rangle$$

(viac kapitola 8.1).



Obr. 7: Intervaly spoľahlivosti pre odhady regresných parametrov

Teraz znovu vypočítame odhad periodickej zložky podľa rovnakého vzorca ako predtým, avšak už bez členov súvisiacich s nevýznamnými parametrami.



Obr. 8: Porovnanie modelu so všetkými parametrami a modelu bez zbytočných parametrov

Ako vidieť z grafu 8, modelové krivky sa líšia minimálne, odobrané parametre

boli naozaj nevýznamné, resp. nepotrebné.

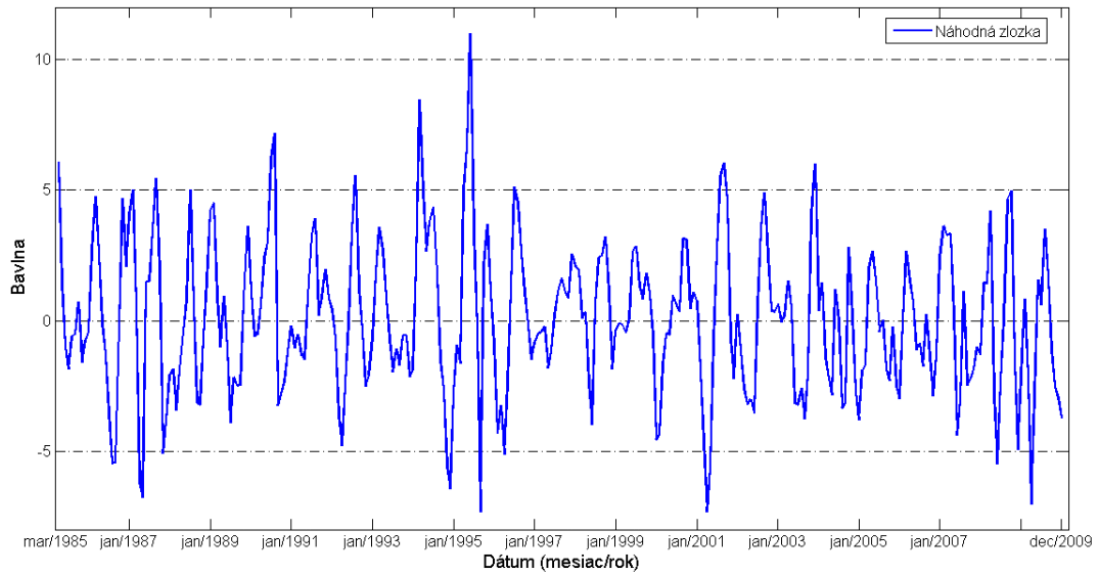
Teraz môžeme od modelu

$$y_t' = P_t + \epsilon_t$$

odpočítať ďalšiu zložku - periodickú  $P_t$ . A ostáva nám model

$$y_t'' = \epsilon_t$$

znázornený na obr. 9.



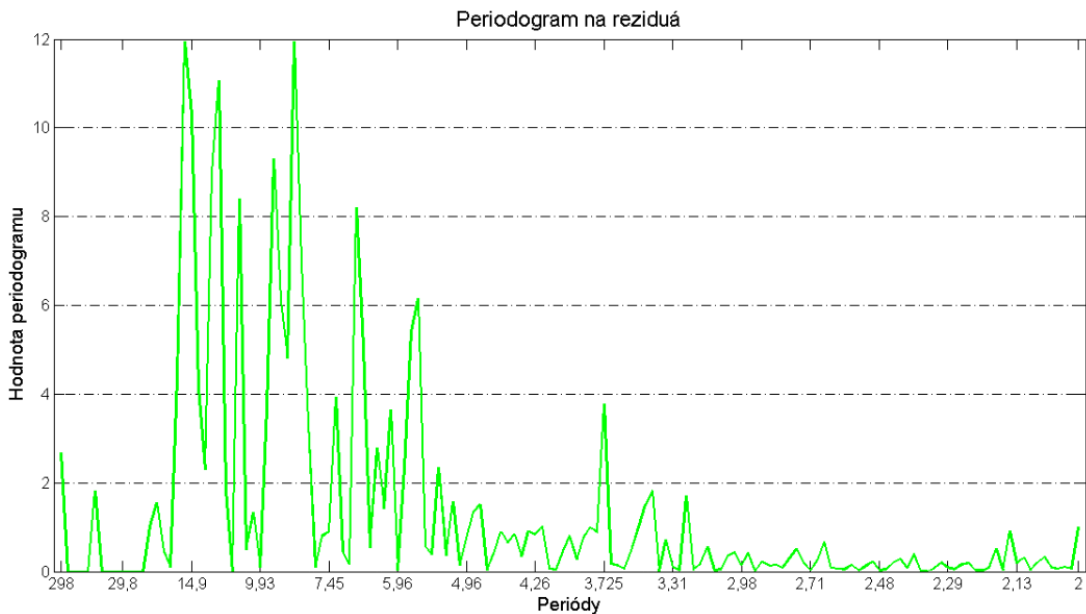
Obr. 9: Náhodná zložka

V tomto modeli musíme otestovať, či sú vplyvy naozaj náhodné a či neobsahuje ešte nejakú dodatočnú nezahrnutú periodickosť. Preto aj na tieto dáta aplikujeme periodogram (obr. 10).

Hodnoty periodogramu si opäť zoradíme zostupne a otestuje prvú, najväčšiu hodnotu Fisherovým testom významnosti. Táto prislúcha perióde dĺžky 15.684. Fisherova štatistika

$$W = \frac{V_1}{V_1 + V_2 + \dots + V_{149}} = 0.058835$$

a pri hladine významnosti  $\alpha = 0.01$  je kritická hodnota  $x = 0.0629$ . Fisherova štatistika je menšia ako kritická hodnota, preto môžeme prehlásiť, že náhodná



Obr. 10: Periodogram náhodnej zložky

zložka je už iba biely šum,  $WN(0, \sigma^2)$ .

Nulovosť strednej hodnoty môžeme overiť jednovýberovým t-testom (viď kapitola 8.3) Budeme testovať nulovú hypotézu

$$H_0 : \mu = 0 \quad \text{proti} \quad H_1 : \mu \neq 0$$

pričom testová štatistika bude

$$T = \frac{\bar{y}_t'' - 0}{S} \sqrt{298} = -7.85e - 015,$$

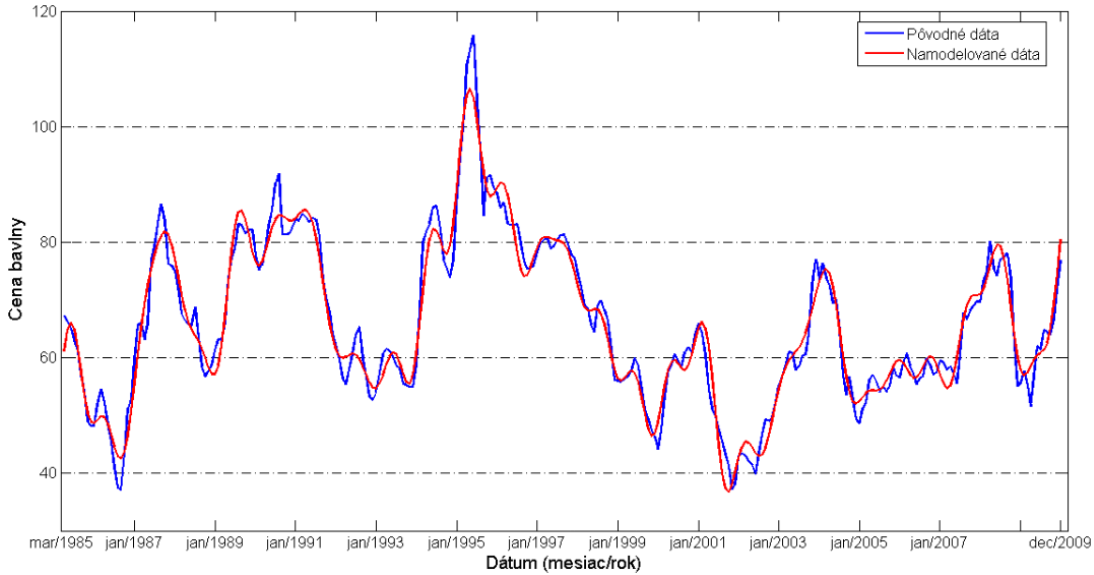
pričom  $\bar{y}_t''$  je priemer všetkých hodnôt časového radu náhodnej zložky. Testová štatistika  $T$  v absolútnej hodnote je menšia ako kvantil  $t_{297}(0.05) = 1.65$  preto nulovú hypotézu nemôžeme zamietnuť za strednú hodnotu teda budeme považovať nulu.

Teraz, keď už máme každú zložku časového radu izolovanú a preskúmanú, môžeme jednotlivé zložky opäť spojiť a vytvoriť model s predpisom

$$\hat{y}_t = \hat{T}_t + \hat{P}_t$$

a porovnať ho s pôvodným časovým radom (obr. 11).





Obr. 11: Grafické porovnanie pôvodných a namodelovaných dát

Modelová funkcia síce veľmi dobre popisuje dáta (index determinácie = 0.9560, reziduálny súčet štvorcov = 2543.22), avšak pokúsime sa model ešte vylepšiť. Modelovú funkciu  $\hat{y}_t = \hat{T}_t + \hat{P}_t$  rozvineme do Taylorovho rozvoja, kde derivácie druhého a vyššieho rádu zanedbáme. Dostávame

$$\hat{y}_t''' = \hat{T}_t(\delta'_0) + \hat{P}_t(\beta'_0) + \left( \frac{\partial \hat{T}_t}{\partial \delta} \Delta\delta + \frac{\partial \hat{P}_t}{\partial \beta} \Delta\beta \right) \Bigg|_{\delta=\delta'_0, \beta=\beta'_0},$$

kde  $\delta'_0 = \hat{\delta}$  a  $\beta'_0 = \hat{\beta}$  sú odhad parametrov získané už skôr v priebehu výpočtu a

$$\Delta\delta = \delta - \delta'_0 \quad \Delta\beta = \beta - \beta'_0.$$

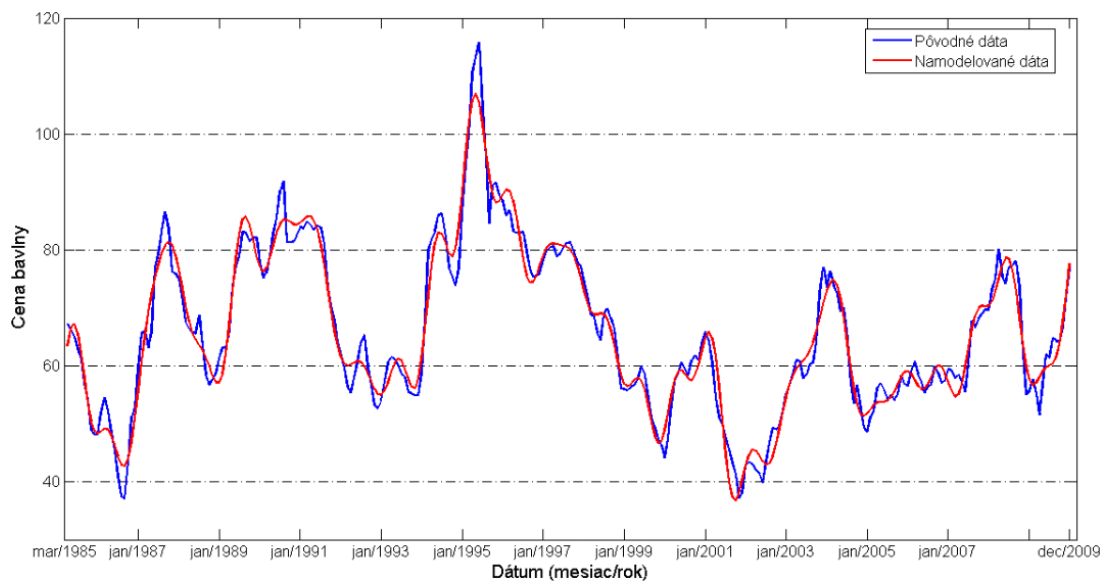
Tento postup budeme opakovať, až kým zmena parametrov  $\delta$  a  $\beta$  bude zanedbateľná. V tabuľke 7 uvádzame hodnoty parametrov po prvej iterácii pomocou Taylorovho rozvoja ako aj hodnoty  $\Delta\delta$  a  $\Delta\beta$ .

Pri druhej iterácii pomocou Taylorovho rozvoja sú hodnoty  $\Delta\delta$  a  $\Delta\beta$  rádovo okolo  $10^{-15}$ , takže ich môžeme zanedbať a za odhady parametrov  $\delta$  a  $\beta$  budeme považovať odhady získané po prvej iterácii.

Po tejto úprave sa nám index determinácie zväčšil na 0.9574 a reziduálny súčet štvorcov zmenšil na 2461.49.

	$\delta'_0, \beta'_0$	$\Delta$	$\delta, \beta$		$\delta'_0, \beta'_0$	$\Delta$	$\delta, \beta$
$\hat{\delta}_0$	65,706	-3,95e-015	65,706	$\hat{\beta}_{16}$	1,032	-0,358	0,674
$\hat{\delta}_1$	-0,034	-0,009	-0,043	$\hat{\beta}_{17}$	2,476	0,022	2,498
$\hat{\delta}_2$	-3.69e-004	-4,02e-005	-4.09e-004	$\hat{\beta}_{18}$	2,506	-0,153	2,353
$\hat{\delta}_3$	1,32e-005	-1,21e-006	1,20e-005	$\hat{\beta}_{19}$	1,693	0,022	1,715
$\hat{\beta}_1$	-6,059	0,058	-6,001	$\hat{\beta}_{20}$	3,024	-0,167	2,857
$\hat{\beta}_2$	3,926	-0,561	3,365	$\hat{\beta}_{21}$	0,686	0,024	0,71
$\hat{\beta}_3$	1,924	0,041	1,965	$\hat{\beta}_{22}$	-2,646	-0,203	-2,849
$\hat{\beta}_4$	-5,454	-0,439	-5,893	$\hat{\beta}_{23}$	-2,263	0,027	-2,236
$\hat{\beta}_5$	-5,411	0,029	-5,382	$\hat{\beta}_{24}$	-0,979	-0,26	-1,239
$\hat{\beta}_7$	3,397	0,106	3,503	$\hat{\beta}_{26}$	2,303	-0,072	2,231
$\hat{\beta}_8$	-3,014	-0,733	-3,747	$\hat{\beta}_{28}$	1,955	-0,06	1,895
$\hat{\beta}_9$	3,936	0,025	3,961	$\hat{\beta}_{30}$	1,788	-0,064	1,724
$\hat{\beta}_{10}$	1,08	-0,228	0,852	$\hat{\beta}_{31}$	1,591	0,020	1,611
$\hat{\beta}_{11}$	1,633	0,023	1,656	$\hat{\beta}_{33}$	-0,95	0,02	-0,93
$\hat{\beta}_{12}$	3,409	-0,183	3,226	$\hat{\beta}_{34}$	-1,322	-0,069	-1,391
$\hat{\beta}_{13}$	0,922	0,022	0,944	$\hat{\beta}_{35}$	-1,501	0,021	-1,48
$\hat{\beta}_{14}$	3,554	-0,141	3,413	$\hat{\beta}_{37}$	1,434	0,021	1,455
$\hat{\beta}_{15}$	3,439	0,033	3,472	$\hat{\beta}_{39}$	1,29	0,021	1,311

Tabuľka 7: Taylorov rozvoj - hodnoty parametrov



Obr. 12: Výsledný model

Výsledný model bude mať predpis :

$$\hat{y}_t = \hat{\delta}_0\phi_0(t) + \hat{\delta}_1\phi_1(t) + \hat{\delta}_2\phi_2(t) + \hat{\delta}_3\phi_3(t) + \sum_{i=1}^{20} \left( \hat{\beta}_{2i-1} \cos \frac{2\pi t}{\tau_i} + \hat{\beta}_{2i} \sin \frac{2\pi t}{\tau_i} \right).$$

Maticovo

$$\hat{y}_t = \mathbf{X}(\hat{\boldsymbol{\delta}}, \hat{\boldsymbol{\beta}})^T.$$

Hodnoty parametrov  $\hat{\delta}_j$  a  $\hat{\beta}_k$  sú uvedené v tabuľke 7, tie ktoré tam uvedené nie sú, sú rovné 0 a hodnoty  $\tau_i$  sú uvedené v tabuľke 5. Odhady všetkých hodnôt  $\hat{y}_t$  sú uvedené v prílohe E.

Kedže model už máme zostavený, môžeme prejsť k predpovedaniu budúceho vývoja. Najprv sa budeme zaoberať bodovými predikciami na 6 mesiacov dopredu. V definícii 9.1 uvádzame vzorec pre výpočet bodových predpovedí.

Vektor  $\hat{\beta}$  sme už odhadli v predchádzajúcich výpočtoch. Ostáva nám určiť vektor  $\mathbf{t}_0$ . Ten bude rozmeru  $36 \times 1$ . Prvé štyri zložky prislúchajú trendovej zložke, ktorá je modelovaná pomocou Čebyševových polynómov. Hodnoty  $\phi_0(t)$  až  $\phi_3(t)$  máme vypočítane iba pre  $t = 1, 2, \dots, 298$ . Keďže ideme určovať hodnotu v časoch  $t = 299, 300, \dots, 304$ , musíme si tieto hodnoty podľa postupu uvedeného v kapitole 6.1.4 dopočítať.

Ďalších 32 zložiek dostaneme dosadením do modelu skrytých periód namiesto  $t$  čas, v ktorom predpoveď počítame. Takto postupujeme pri všetkých šiestich bodových predpovediach.

t	Dátum	Skutočná hodnota	Bodová predpoveď
299	január 2010	77.40	81.713
300	február 2010	80.04	84.175
301	marec 2010	85.79	84.347
302	apríl 2010	88.09	82.16
303	máj 2010	90.07	78.131
304	jún 2010	91.68	73.198

Tabuľka 8: Bodové predpovede ČR ceny bavlny na najbližších 6 mesiacov

V tabuľke 8 sú uvedené bodové predpovede do júla 2010 aj skutočné ceny bavlny. Vidíme, že bodové predpovede úplne nezodpovedajú skutočným hodnotám,

preto si uvedieme aj 95% intervaly spoľahlivosti pre predikcie. Tie budú tvaru

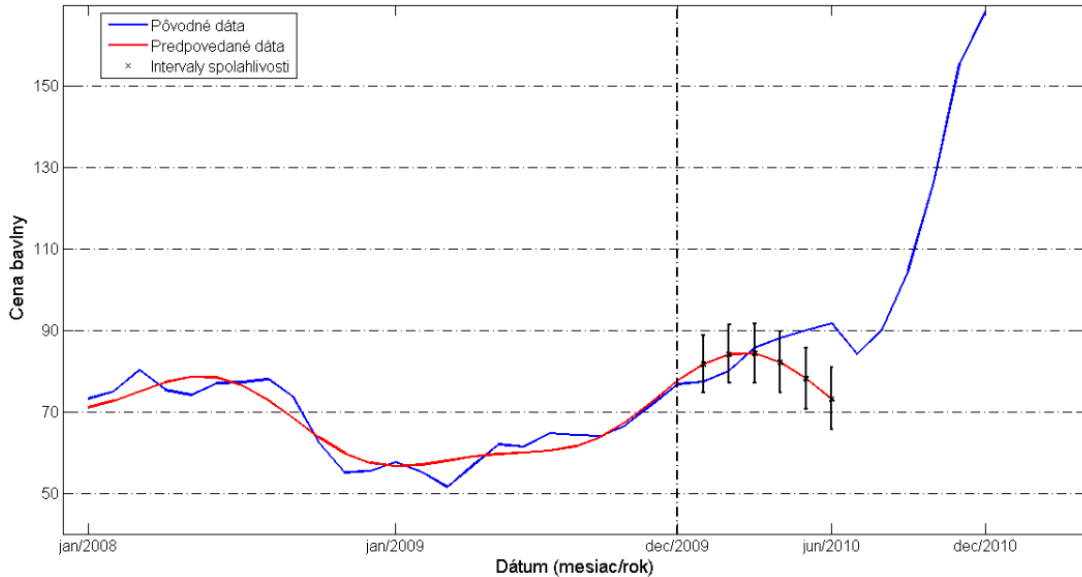
$$\left\langle \hat{y}_0 - t_{262}(0.975)\hat{\sigma}\sqrt{(1 + t_0^T(\mathbf{X}^T\mathbf{X})^{-1}t_0)}; \hat{y}_0 + t_{262}(0.975)\hat{\sigma}\sqrt{(1 + t_0^T(\mathbf{X}^T\mathbf{X})^{-1}t_0)} \right\rangle$$

kde  $\hat{y}_0$  sú bodové predpovede zostrojené v budúcom čase  $t = 299, \dots, 304$ .

t	Dátum	Skutočná hodnota	Interaval spoľahlivosti
299	január 2010	77.40	< 74.956; 88.47 >
300	február 2010	80.04	< 77.228; 91.123 >
301	marec 2010	85.79	< 77.231; 91.462 >
302	apríl 2010	88.09	< 74.915; 89.405 >
303	máj 2010	90.07	< 70.774; 85.487 >
304	jún 2010	91.68	< 65.723; 80.673 >

Tabuľka 9: Intervaly spoľahlivosti ČR ceny bavlny na najbližších 6 mesiacov

Intervaly spoľahlivosti zostrojené pre prvé štyri predikčné časové okamžiky pokrývajú skutočnú hodnotu (tab. 9, obr. 13). Pre prvé tri predikčné obdobia dokonca platí, že rastie cena bavlny a rastie aj predikovaná hodnota. Ak by sme teda podľa tohoto modelu investovali, máme v prvých mesiacoch šancu na zisk. Na grafe ktorý znázorňuje budúci vývoj ceny bavlny (obr. 13) však vidieť aj to, že sa cena v priebehu roka zvýšila zo \$75 na skoro \$170. V priebehu 25 rokov počas ktorých sme tento časový rad sledovali sa cena pohybovala iba v rozmedzí \$35 až \$120 dolárov. Preto je tento vývoj veľmi pozoruhodný a veľmi ťažko zachytiteľný.



Obr. 13: Predikcie na 6 mesiacov aj s intervalmi spoľahlivosti

## 10.2 Časový rad ceny kakaových bôbov na burze

Rovnaký postup sme sa rozhodli vyskúšať aj na časovom rade ceny kakaových bôbov. Cena tejto komodity je uvedená v amerických dolároch za tonu. Opäť sme si vzali rovnaký počet údajov (298), taktiež od marca 1985 do decembra 2009 dostupné z [27]. Všetky sú uvedené v prílohe E a graficky zobrazené na obr. 14. Trendovú funkciu si znovu určíme pomocou Čebyševovej metódy. Vytvorili sme si 5 modelov rovnako ako v prípade hľadania trendovej funkcie ceny bavlny. Prvý model tvaru

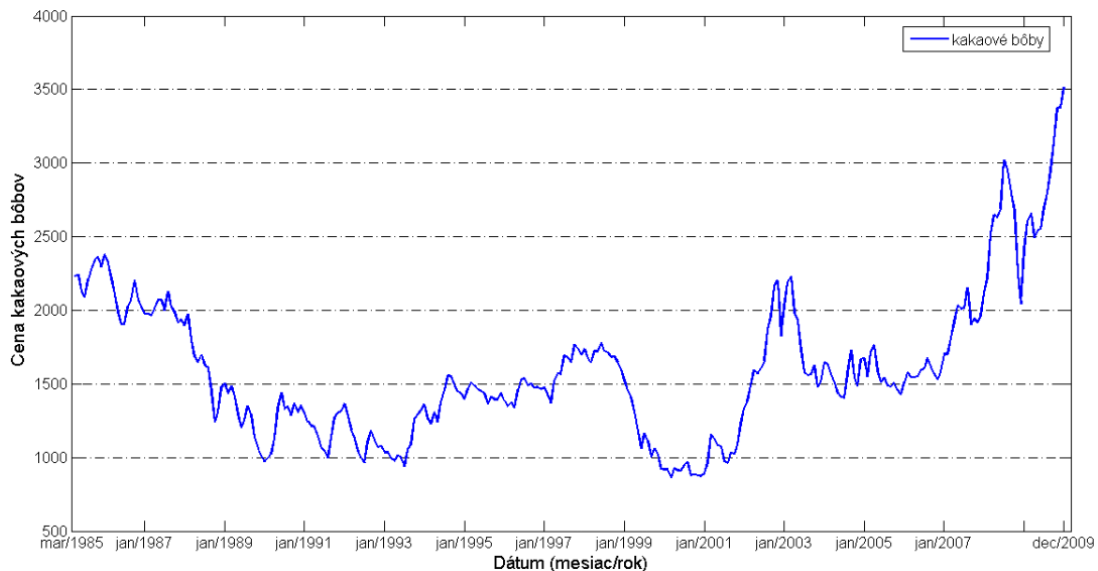
$$T_t = \phi_0(t)\delta_0, \quad t = 1, 2, \dots, 298$$

nazveme modelom nultého stupňa, až piaty model

$$T_t = \phi_0(t)\delta_0 + \phi_1(t)\delta_1 + \phi_2(t)\delta_2 + \phi_3(t)\delta_3 + \phi_4(t)\delta_4, \quad t = 1, 2, \dots, 298$$

nazveme model štvrtého stupňa.

Pri všetkých týchto modeloch si porovnáme reziduálny súčet štvorcov, index determinácie, odhad rozptylu  $\hat{\sigma}^2$  a taktiež otestujeme parametre  $\delta_i$  testom významnosti. Získané údaje sú uvedené v tabuľkách 10 a 11. Odhady všetkých trendových funkcií sú graficky znázornené v prílohe B.



Obr. 14: Vývoj ceny kakaových bôbov od marca 1985 do decembra 2009

Stupeň modelu	Index determ. - $R^2$	Reziduálny súčet štvorcov	$\hat{\sigma}^2$
0.	0	7.3238e+007	246593
1.	0.071	6.8036 e+007	229850
2.	0.619	2.7903 e+007	94587
3.	0.6205	2.7795 e+007	94540
4.	0.7328	1.9569 e+007	66788

Tabuľka 10: Porovnanie modelov

Z tabuliek 10 a 11 vyplýva, že na popis trendovej funkcie nám stačí použiť trendovú funkcie 2. stupňa. Aj keď pri testovaní významnosti parametra  $\hat{\delta}_4$  v modely 4. stupňa tento vyšiel významný, nejedná sa už o štatistickú ale skôr o technickú významnosť. Preto pre odhad  $\hat{T}_t$  bude platiť

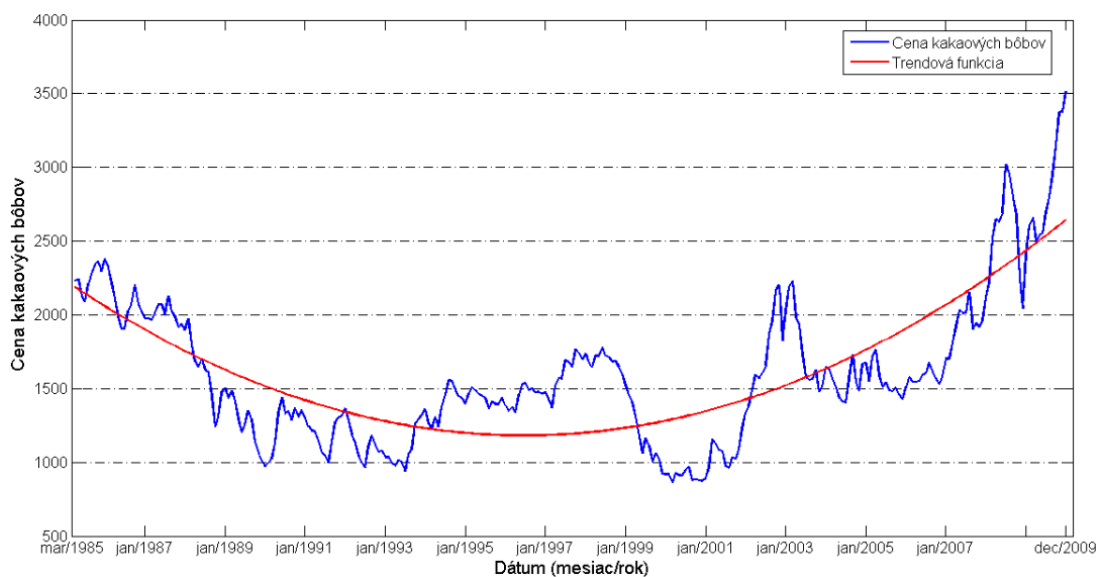
$$\hat{T}_t = 1602.4\phi_0(t) + 1.536\phi_1(t) + 0.055\phi_2(t), \quad t = 1, 2, \dots, 298.$$

Z obr. 15 vidíme, že zvolená funkcia správne charakterizuje vývoj ceny kakaových bôbov.

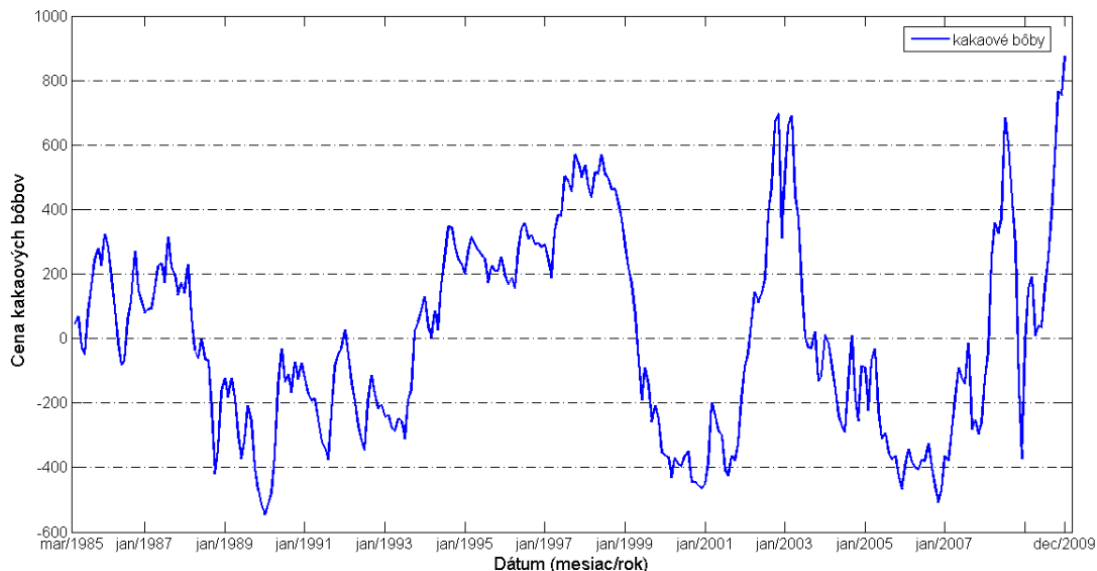
Hodnoty parametrov		Hodnota testovacej štatistiky T / rozhodnutie o zamietnutí				
		Model 0.stupňa	Model 1.stupňa	Model 2.stupňa	Model 3.stupňa	Model 4.stupňa
$\hat{\delta}_0$	1602.4	55.7045 /zam.	57.6975 /zam.	89.9424 /zam.	89.9648 /zam.	107.0362 /zam.
$\hat{\delta}_1$	1.5359		4.7574 /zam.	7.4162 /zam.	7.4180 /zam.	8.8256 /zam.
$\hat{\delta}_2$	0.055444			20.5984 /zam.	20.6035 /zam.	24.5132 /zam.
$\hat{\delta}_3$	3.8147e-005				1.0710 /nezam.	1.2743 /nezam.
$\hat{\delta}_4$	4.425e-006					11.0979 / zam.

Tabuľka 11: Hodnoty testovacej štatistiky T

V ďalšom kroku od pôvodného časového radu  $y_t$  odpočítame odhad trendovej zložky  $\hat{T}_t$  a dostávame časový rad (obr. 16), ktorý obsahuje už iba periodickú a náhodnú zložku, ktoré budeme ďalej skúmať.



Obr. 15: Odhad trendovej funkcie



Obr. 16: Cena kakaových bôbov po odpočítaní trendovej funkcie

Na skúmanie peridickej zložky použijeme periodogram, kde uvidíme, či náš časový rad obsahuje skryté periodicity alebo nie. Pre výpočet použijeme vzorec z lemy 6.1 a výsledné hodnoty zakreslíme do grafu (obr. 17) a uvádzame ich v prílohe E.

Vidíme, že periodogram obsahuje určité vrcholy, ktoré sú výrazne vyššie ako ostatné, ale nevieme, ktoré do nášho modelu ešte zahrnúť a ktoré už nie, preto ich postupne otestujeme Fisherovým testom. Hodnoty periodogramu si zoradíme od najväčšej po najmenšiu a začneme s testovaním najväčšej z nich. Vypočítame si hodnotu Fisherovej štatistiky

$$W = \frac{V_1}{\sum_{i=1}^{149} V_i} = 0.2588$$

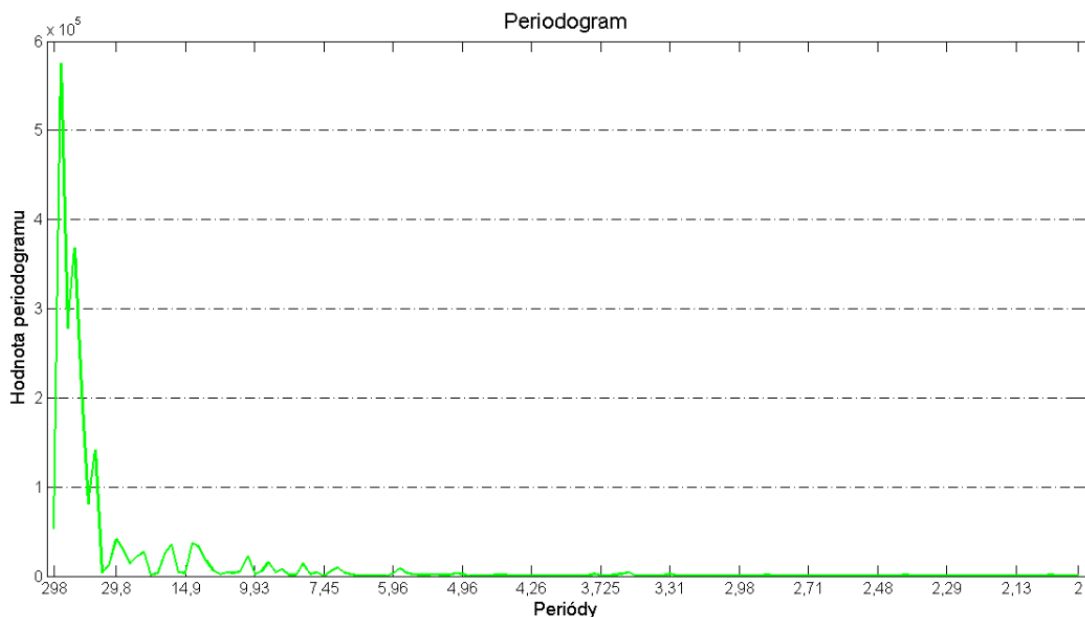
pre prvú, najväčšiu hodnotu periodogramu  $V_1$  a túto porovnáme s aproximovanou hodnotou kritickej hodnoty Fisherovho testu  $x = 0.0629$ . Hodnota Fisherovej štatistiky je väčšia ako kritická hodnota, preto periódu prislúchajúcu najvyššej hodnote periodogramu prijímame ako významnú.

Rovnako otestujeme aj ďalšie hodnoty, až kým nenarazíme na prvú nevýznamnú. V tabuľke v prílohe C uvádzame hodnoty Fisherovej štatistiky a prislúchajúcej kritickej hodnoty až po prvú hodnotu periodogramu, ktorú zamietneme.

Fisherovým testom na hladine významnosti 99% sa ukázalo, že náš časový rad



obsahuje sedemnášť významných periód s dĺžkami uvedenými v tabuľke 12.



Obr. 17: Periodogram časového radu ceny kakaových bôbov

$\tau_1$	149	$\tau_2$	74.5	$\tau_3$	99.33	$\tau_4$	59.60
$\tau_5$	42.57	$\tau_6$	49.67	$\tau_7$	298	$\tau_8$	29.80
$\tau_9$	14.19	$\tau_{10}$	16.56	$\tau_{11}$	13.55	$\tau_{12}$	27.09
$\tau_{13}$	21.29	$\tau_{14}$	17.53	$\tau_{15}$	10.28	$\tau_{16}$	22.92
$\tau_{17}$	12.96						

Tabuľka 12: Hodnoty parametrov  $\tau_i$  ČR ceny kakaových bôbov

Odhad modelu periodickej zložky teraz môžeme zapísať

$$\hat{P}_t = \sum_{i=1}^{17} \left( \hat{\beta}_{2i-1} \cos \frac{2\pi t}{\tau_i} + \hat{\beta}_{2i} \sin \frac{2\pi t}{\tau_i} \right) = \{\mathbf{X}\}_t \hat{\boldsymbol{\beta}}, \quad t = 1, 2, \dots, 298,$$

kde  $\hat{\beta}_i, i = 1, 2, \dots, 34$  sú odhady získané metódou najmenších štvorcov.

Ostáva nám ešte otestovať významnosť regresných parametrov  $\hat{\beta}_1, \dots, \hat{\beta}_{34}$ . Budeme testovať hypotézu

$$H_0 : \hat{\beta}_i = 0 \quad \text{proti} \quad H_1 : \hat{\beta}_i \neq 0$$

pomocou testovacej štatistiky

$$T = \frac{\beta_i - 0}{\hat{\sigma} \sqrt{v_{ii}}},$$

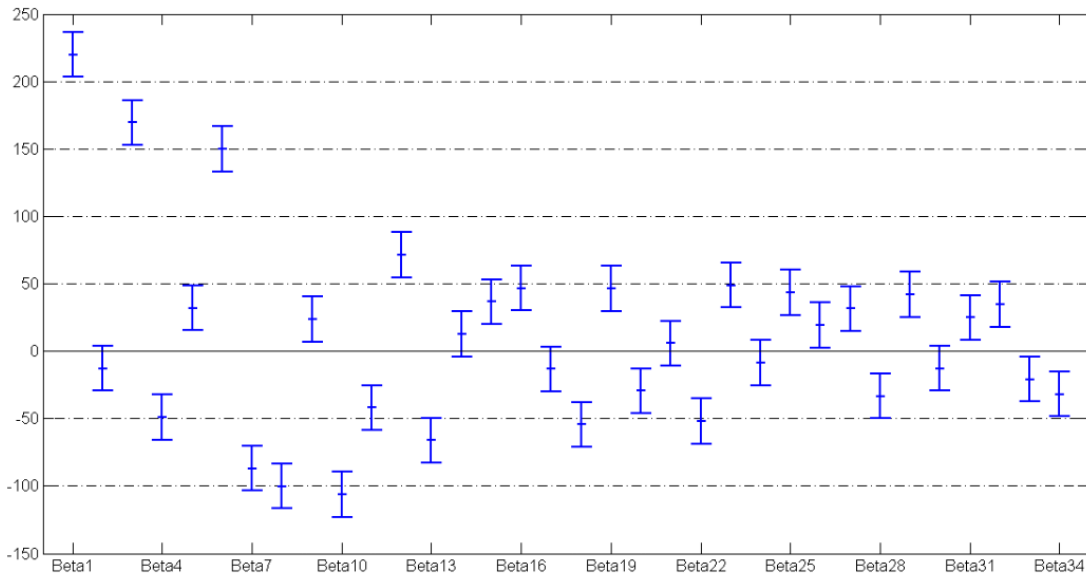
kde  $v_{ii}$  sú diagonálne prvky matice  $(\mathbf{X}^T \mathbf{X})^{-1}$  a  $\hat{\sigma}^2 = (\mathbf{y}_t' - \hat{\mathbf{P}}_t)^T (\mathbf{y}_t' - \hat{\mathbf{P}}_t) / (298 - 34) = 10707$ ,  $y_t'$  je pôvodný časový rad po odpočítaní trendovej zložky.

Testovaciu štatistiku  $T$  porovnáваме s kritickou hodnotou tohto testu  $t_{0.975}(264) = 1.969$ . Ak je  $T$  menšie ako kritická hodnota, nulovú hypotézu nezamietame a parameter  $\beta_i$ , pre ktorý takáto nerovnosť platí považujeme za nevýznamný a v modely ho zanedbávame.

Parameter	T	Rozhodnutie	Parameter	T	Rozhodnutie
$\beta_1$	25.9279	zamietame	$\beta_{18}$	-6.4307	zamietame
$\beta_2$	-1.5050	nezamietame	$\beta_{19}$	5.4463	zamietame
$\beta_3$	19.9704	zamietame	$\beta_{20}$	-3.4794	zamietame
$\beta_4$	-5.7948	zamietame	$\beta_{21}$	0.6675	nezamietame
$\beta_5$	3.7767	zamietame	$\beta_{22}$	-6.1276	zamietame
$\beta_6$	17.6808	zamietame	$\beta_{23}$	5.7643	zamietame
$\beta_7$	-10.2589	zamietame	$\beta_{24}$	-1.0347	nezamietame
$\beta_8$	-11.8137	zamietame	$\beta_{25}$	5.1275	zamietame
$\beta_9$	2.7632	zamietame	$\beta_{26}$	2.2551	zamietame
$\beta_{10}$	-12.5632	zamietame	$\beta_{27}$	3.7054	zamietame
$\beta_{11}$	-4.9471	zamietame	$\beta_{28}$	-3.9374	zamietame
$\beta_{12}$	8.4280	zamietame	$\beta_{29}$	4.9294	zamietame
$\beta_{13}$	-7.8053	zamietame	$\beta_{30}$	-1.5163	nezamietame
$\beta_{14}$	1.4712	nezamietame	$\beta_{31}$	2.9261	zamietame
$\beta_{15}$	4.3130	zamietame	$\beta_{32}$	4.0924	zamietame
$\beta_{16}$	5.5006	zamietame	$\beta_{33}$	-2.4651	zamietame
$\beta_{17}$	-1.5852	nezamietame	$\beta_{34}$	-3.7597	zamietame

Tabuľka 13: Testovanie významnosti parametrov modelu skrytých periód

Z tabuľky 13 vidíme, že náš model obsahuje šesť nevýznamných parametrov, ktoré môžeme vynechať. To potvrdzujú aj intervaly spoľahlivosti zostrojené podľa vzorca 8, ktoré sú graficky znázornené na obr. 18. Ak interval pre určitý parameter obsahuje nulu, tento parameter je nevýznamný. Po otestovaní všetkých parametrov a vylúčení nevýznamných môžeme konečne napísať predpis pre odhad



Obr. 18: Intervaly spoľahlivosti pre parametre  $\beta_i$

periodickej zložky pomocou modelu skrytých periód

$$\hat{P}_t = \sum_{i=1}^{17} \left( \hat{\beta}_{2i-1} \cos \frac{2\pi t}{\tau_i} + \hat{\beta}_{2i} \sin \frac{2\pi t}{\tau_i} \right), \quad t = 1, 2, \dots, 298,$$

kde sme odhady  $\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_{34}$  získali metódou najmenších štvorcov a nevýznamné parametre sme nahradili nulou, resp. z modelu vylúčili a parametre  $\tau_1$  až  $\tau_{17}$  sú uvedené v tabuľke 12 a vypočítali sme ich pomocou periodogramu a Fisherovho testu. Graf charakterizujúci periodickú zložku je znázornený na obr. 19.

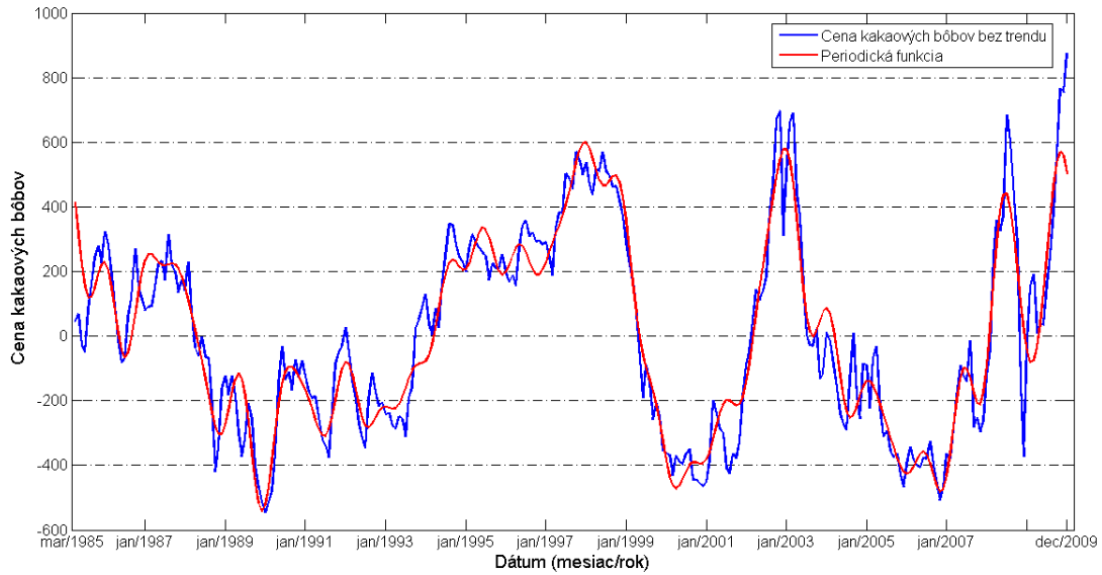
Z pôvodného modelu

$$y_t = T_t + P_t + \epsilon_t, \quad t = 1, 2, \dots, 298$$

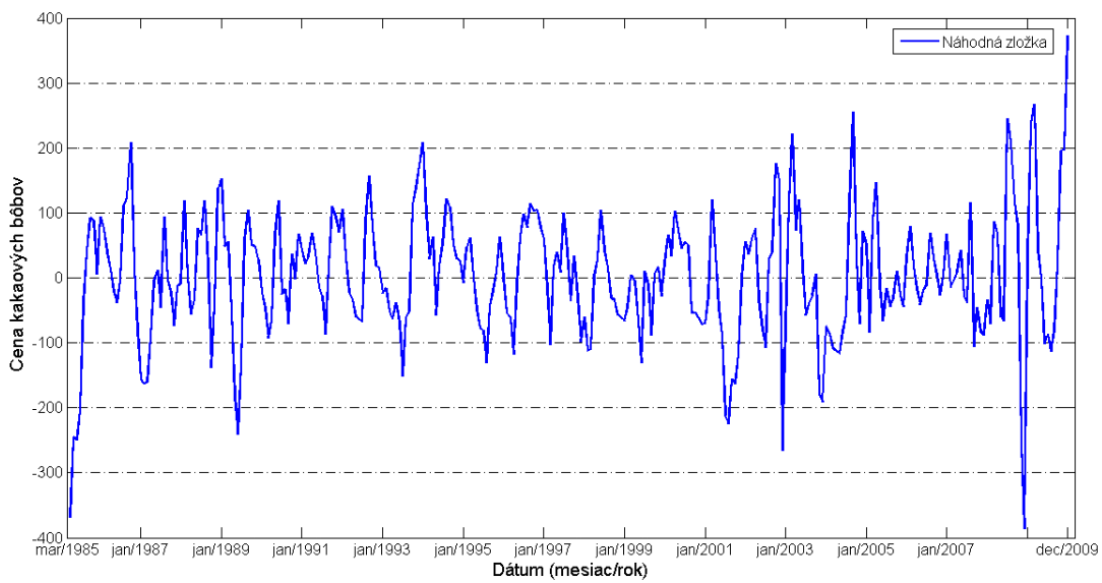
máme odhadnutú trendovú aj periodickú zložku, ostáva nám teda už iba preskúmať náhodnú zložku  $\epsilon_t$  zobrazenú na obr. 20.

V prvom rade sa pokúsime otestovať, či náhodou časový rad charakterizujúci náhodnú zložku neobsahuje ešte nejakú periodicitu. Na dáta popisujúce náhodnú zložku aplikujeme periodogram vysvetlený v kapitole 6.2.1. Graficky je periodogram náhodnej zložky vykreslený na obr. 21.

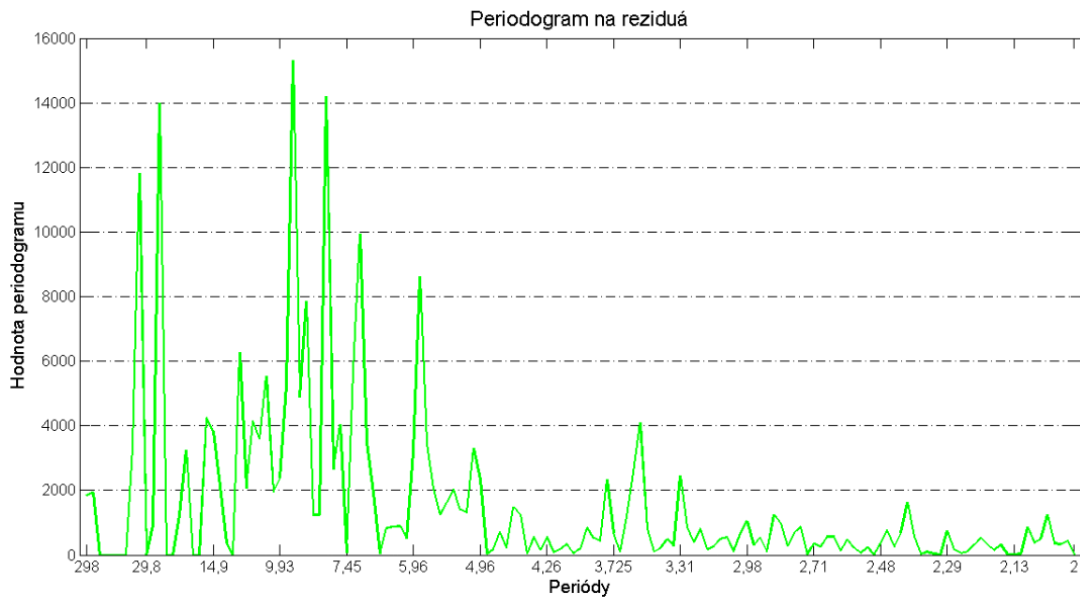
Vidíme, že najväčšiu hodnotu dosahuje periodogram pre periódu dĺžky 9.3125 a to okolo 15500, potom pre periódu dĺžky 8.054, 24.83, atď. Či sú tieto hodnoty



Obr. 19: Odhad periodickej funkcie ceny kakaových bôbov



Obr. 20: Náhodná zložka ČR ceny kakaových bôbov



Obr. 21: Periodogram náhodnej zložky ČR ceny kakaových bôbov

výrazne väčšie ako ostatné však musíme otestovať Fisherovým testom (viď kapitola 6.2.2). Hodnoty Fisherovej štatistiky a kritickej hodnoty pre prvých niekoľko hodnôt periodogramu, až po prvú, ktorú test označí za nevýznamnú sú uvedené v tabuľke 14. Z tabuľky vyplýva, že časový rad náhodnej zložky obsahuje ešte tri

	perióda	Fisherová štatistika	Kritická hodnota
1	9.3125	0.065356	0.062864
2	8.0541	0.064838	0.063235
3	24.833	0.068378	0.06361
4	33.111	0.06195	0.06399

Tabuľka 14: Hodnoty fisherovej štatistiky a kritickej hodnoty pre prvé štyri najväčšie hodnoty periodogramu náhodnej zložky ČR kakaových bôbov

periodicity, ktoré musíme zahrnúť do periodickej zložky. Jej odhad sa nám teda zmení na

$$\hat{P}_t = \sum_{i=1}^{20} \left( \hat{\beta}_{2i-1} \cos \frac{2\pi t}{\tau_i} + \hat{\beta}_{2i} \sin \frac{2\pi t}{\tau_i} \right) = \{\mathbf{X}\}_t \hat{\beta},$$

kde prvých 17 parametrov  $\tau_i$  sme získali z periodogramu časovej rady ceny kakaových bôbov očistenej od trendu a posledné tri z periodogramu náhodnej zložky.

Parametre  $\beta_i, i = 1, 2, \dots, 40$  sme opäť získali metódou najmenších štvorcov a otestovali sme ich testom významnosti (viď kapitola 8.2). Ako nevýznamný sa nám okrem parametrov, ktoré sa ukázali ako nevýznamné už pri testovaní periodickej zložky, ukázal aj parameter  $\hat{\beta}_{40}$ , ktorý sme taktiež z modelu vylúčili.

Náhodná zložka nášho nového modelu už neobsahuje skryté periodicity a jej hodnoty sa pohybujú v okolí 0, pôjde už iba o biely šum  $WN(0, \sigma^2)$ . Strednú hodnotu si môžeme otestovať jednovýberovým t-testom (viď kapitola 8.3), kde testová štatistika  $T = 1.7713e - 016$  a kritická hodnota  $t_{297}(0.05) = 1.65$ , takže hypotézu  $H_0 : \mu = 0$  nezamietame.

Potom ako sme odhadli všetky zložky modelu, môžeme vytvoriť konečný odhad časového radu ceny kakaových bôbov. Ten bude tvaru

$$\hat{y}_t = \hat{T}_t + \hat{P}_t.$$

Použili sme aditívny tvar nakoľko lepšie kopíruje nami skúmané dáta.

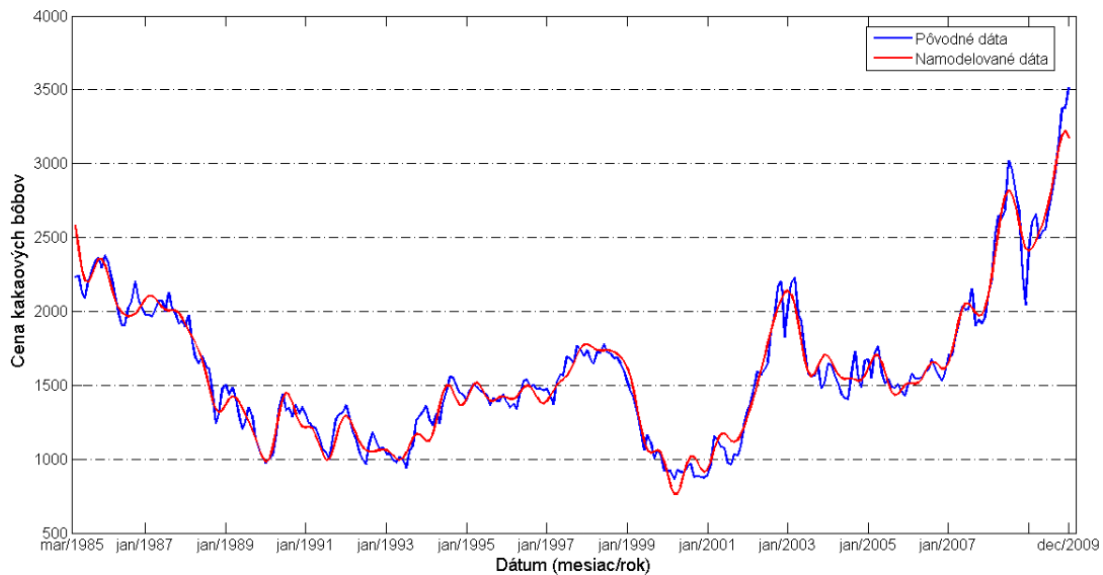
Všetky parametre a premenné sme už počas výpočtu odhadli, pokúsime sa však celkový model ešte vylepšiť obdobne ako v prípade časového radu ceny bavlny. Odhadujúcu funkciu si rozvineme do Taylorovho rozvoja okolo približného bodu  $\delta_0 = \hat{\delta}$  pre odhad trendovej funkcie a  $\beta_0 = \hat{\beta}$  pre odhad periodickej funkcie.

$$\hat{y}_t' = \hat{T}_t(\delta_0) + \hat{P}_t(\beta_0) + \left( \frac{\partial \hat{T}_t}{\partial \delta} \Delta\delta + \frac{\partial \hat{P}_t}{\partial \beta} \Delta\beta \right) \Bigg|_{\delta=\delta_0, \beta=\beta_0},$$

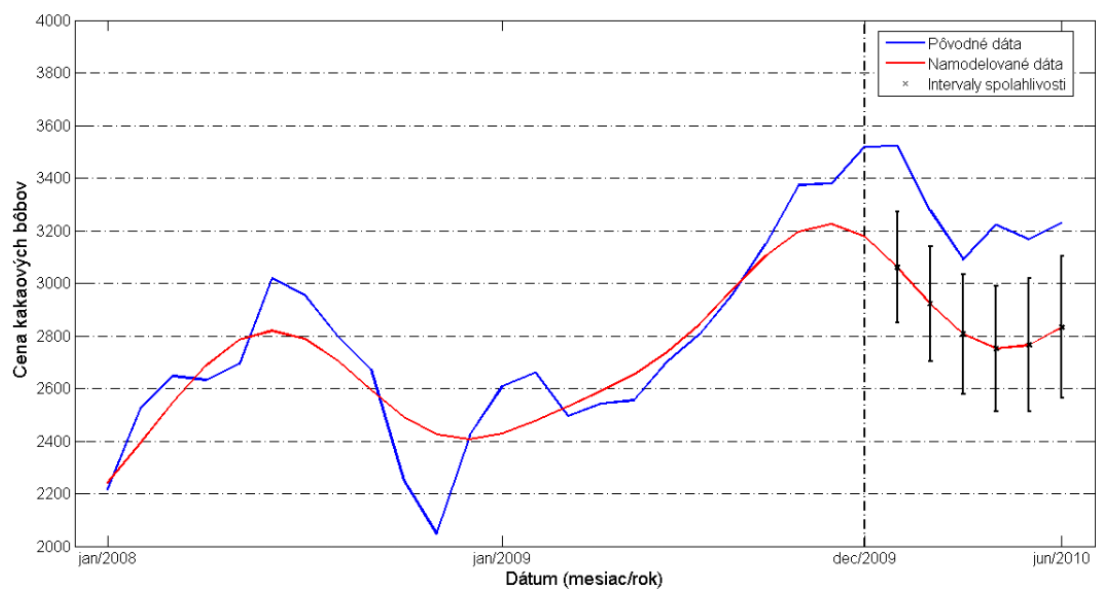
Nové parametre získame zo vzťahov  $\delta = \Delta\delta + \delta_0$  a  $\beta = \Delta\beta + \beta_0$ .

Index determinácie pôvodného modelu je 0.96728 a reziduálny súčet štvorcov 2.3962e+006. Po prvej aproximácii Taylorovým rozvojom sa index determinácie zmenil minimálne na 0.96734, ale RSS sa zmenšil na 2.3913e+006 a teda tento model budeme považovať za lepší. Po druhej aproximácii sú odhady  $\Delta\delta$  a  $\Delta\beta$  zanedbateľne malé, preto predchádzajúci model budeme považovať za odhad pôvodnej časovej rady. Graficky je odhad zobrazený na obr. 22, výsledné hodnoty parametrov sú uvedené v prílohe D a výsledné číselné namodelované hodnoty v prílohe E.

Po zostavení modelu môžeme prejsť ku konštrukcii bodových a intervalových predpovedí bližšie popísaných v kapitole 9. V tabuľke 15 uvádzame hodnoty bodových aj intervalových predpovedí na najbližších 6 mesiacov.



Obr. 22: Výsledný model charakterizujúci časový rad ceny kakaových bôbov



Obr. 23: Predikcie na najbližších 6 mesiacov pre časový rad ceny kakaových bôbov

t	Dátum	Skutočnosť	Bodová predpoveď	Predikčný interval
299	jan 2010	3522.10	3060.36	< 2848.05; 3272.66 >
300	feb 2010	3276.56	2921.50	< 2703.14; 3139.87 >
301	mar 2010	3089.65	2806.84	< 2579.43; 3034.24 >
302	apr 2010	3221.23	2751.16	< 2511.86; 2990.45 >
303	máj 2010	3165.75	2765.02	< 2511.49; 3018.56 >
304	jún 2010	3230.83	2831.95	< 2562.10; 3101.79 >

Tabuľka 15: Predikcie a predikčné intervaly od januára 2010 do júna 2010

Z tabuľky 15 a obr. 23 pre predikcie vidno, že 95% intervaly spoľahlivosti nepokrývajú skutočnú hodnotu ani pre jeden predikovaný časový okamžik, avšak graf v inej cenovej hladine pekne kopíruje skutočný vývoj ceny, takže zachytáva jej zmenu. Niekedy aj tieto informácie investorovi stačia, aby sa správne rozhodol.

Všetky výpočty boli zrealizované v programe MATLAB. V prílohe E sa nachádzajú dva súbory (jeden pre časový rad ceny bavlny - `bavlna.m`, druhý pre ČR ceny kakaových bôbov - `kakao_boby.m`), ktoré sme vytvorili a ktoré po spustení vypočítajú hodnoty použitých premenných a parametrov (vrátane tých, ktoré v práci kôli rozsahu nie sú uvedené) a taktiež vykreslia grafy pre názornejšiu predstavu.



## Záver

Cieľom tejto práce bolo štatisticky preskúmať vybrané časové rady cien komodít na burze, vytvoriť model, ktorý by tieto časové rady čo najvernejšie popisoval a pomocou neho predikovať vývoj na najbližších 6 mesiacov.

Vybrali sme si dva časové rady - ČR ceny bavlny a ČR ceny kakaových bôbov, oba od marca 1985 do decembra 2009, pričom šlo o mesačné dáta. V ekonomickej časti sme uviedli akýsi úvod, ktorý by mal čitateľovi pomôcť pochopiť obchodovanie s komoditami.

Čo sa týka matematickej časti, prvoradou podmienkou predtým ako sme k modelovaniu mohli pristúpiť, bolo naštudovanie potrebnej teórie. Jednalo sa najmä o teóriu časových radov a to klasický dekompozičný prístup a lineárnu regresiu. Následne sme mohli v praktickej časti nami získané teoretické poznatky aplikovať.

Pri oboch časových radoch sme si najprv určili trend vývoja, ktorý bol pri ČR ceny bavlny charakterizovaný polynómom 3. stupňa, pri druhom časovom rade iba polynómom 2. stupňa, oba zostavené pomocou polynomickej regresie.

Následne sme od pôvodného časového radu odpočítali odhad trendovej zložky a analyzovali sme zložku periodickú. Pri oboch časových radoch sme očakávali, že v nich bude periodogramom a následne Fisherovým testom označená za významnú perióda dĺžky približne 12 mesiacov, nakoľko ide o poľnohospodárske produkty, ktoré majú ročný vegetačný cyklus. Tento predpoklad sa nám potvrdil pri oboch komoditách, aj keď táto perióda nebola označená za najvýznamnejšiu.

Po namodelovaní periodickej zložky sme aj túto od modelu odčítali a ostala nám zložka náhodna, kde sme skúmali, či náhodou neobsahuje ešte nejakú skrytú periodicitu. Tú obsahovala iba náhodná zložka časového radu ceny kakaových bôbov.

Čo sa týka modelovania časového radu v minulosti, dosiahli sme veľmi dobrých výsledkov. Pri modeli ČR ceny bavlny je reziduálny súčet štvorcov  $RSS = 2461.49$  a index determinácie 0.9574, čo je veľmi blízko jednej, takže model veľmi presne popisuje dáta a pri časovom rade ceny kakaových bôbov je index determinácie dokonca 0.9673, čo je ešte lepší výsledok a  $RSS = 2.3913e + 006$ .

Po preskúmaní minulého vývoja sme zostavili predikcie aj s predikčnými intervalmi. V predpovedaní budúcnosti už modely neboli príliš presné a pri ČR ceny bavlny nám 95% konfidenčné intervaly pokryli iba prvé 4 skutočné hodnoty, pri

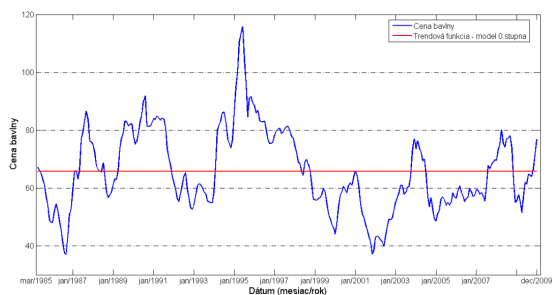
ČR ceny kakaových bôbov dokonca ani jednu. Tento jav môže byť zapríčinený mnohými nepredvídateľnými vplyvmi, ktoré sú takýmto modelom ťažko zachytiteľné, napr. pri bavlně vývoju ceny výrazne napomohlo nepriaznivé počasie a slabšia úroda v Ázii a tak počas roku 2010 bavlna zdvojnásobila svoju cenu, viac [24], kakao zase v júli roku 2010 lámalo rekordy potom ako istý Brit kúpil 240 tisíc ton kakaa v cene miliarda dolárov [25].

Lepšie predikcie a popis modelu by sme možno získali zvolením iného prístupu k modelovaniu týchto časových radov a to napríklad použitím Box-Jenkinsonovej metodológie alebo metódy exponenciálneho vyrovnávania, čo však z dôvodu rozsahu práce nebolo možné vyskúšať.

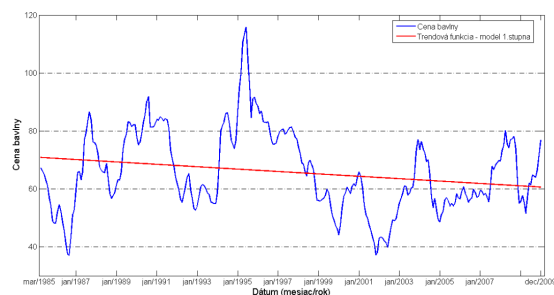
Dúfame, že táto diplomová práca splnila svoj cieľ a objasnila modelovanie v oblasti časových radov a rozšírili čitateľove obzory aj o oblasť obchodovania komodít.

# Prílohy

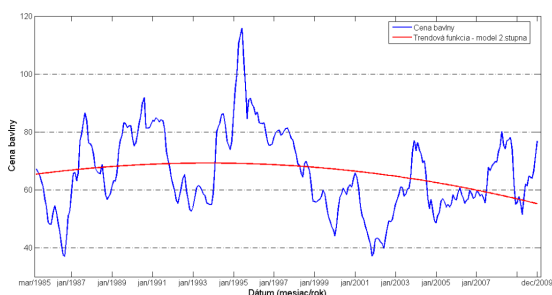
## A Trendové funkcie časového radu ceny bavlny



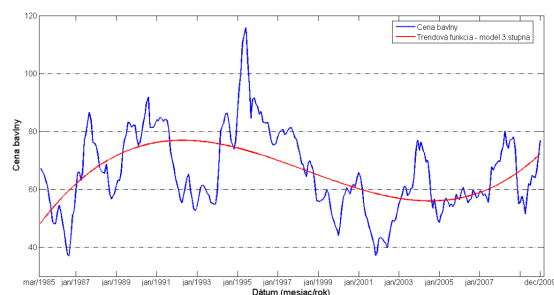
Obr. 24: Odhad trendovej funkcie modelom 0. stupňa



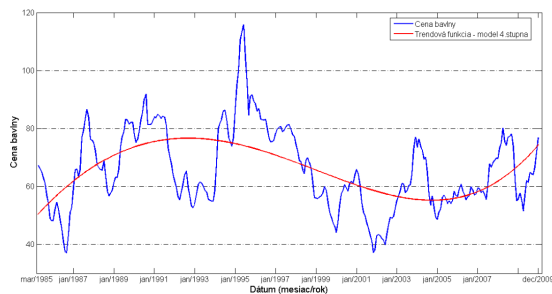
Obr. 25: Odhad trendovej funkcie modelom 1. stupňa



Obr. 26: Odhad trendovej funkcie modelom 2. stupňa

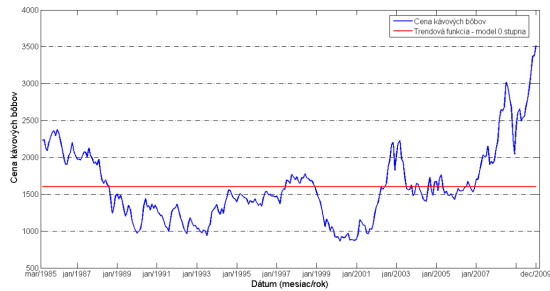


Obr. 27: Odhad trendovej funkcie modelom 3. stupňa

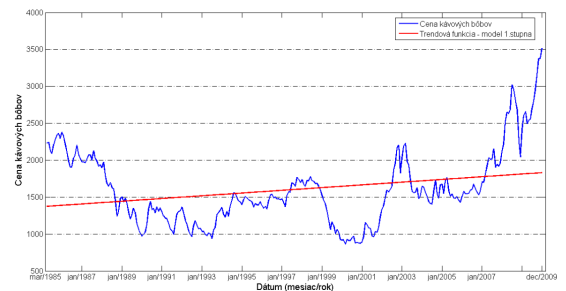


Obr. 28: Odhad trendovej funkcie modelom 4. stupňa

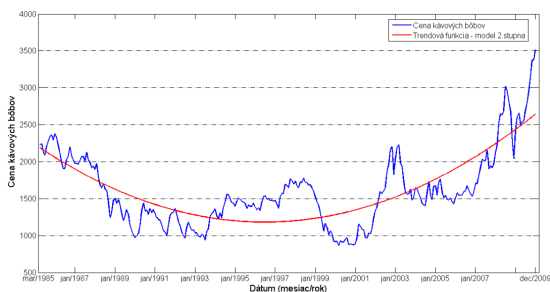
## B Trendové funkcie časového radu ceny kakaových bôbov



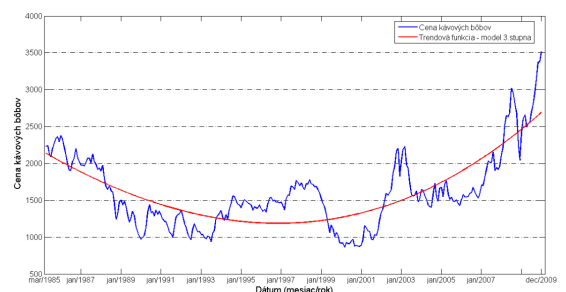
Obr. 29: Odhad trendovej funkcie ČR kakaových bôbov modelom 0. stupňa



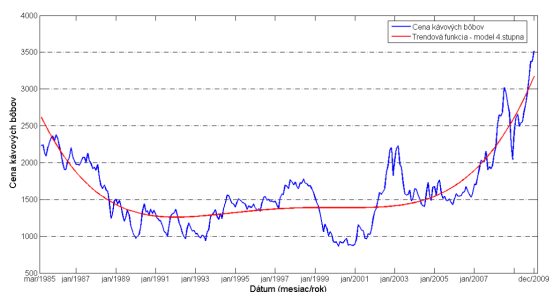
Obr. 30: Odhad trendovej funkcie ČR kakaových bôbov modelom 1. stupňa



Obr. 31: Odhad trendovej funkcie ČR kakaových bôbov modelom 2. stupňa



Obr. 32: Odhad trendovej funkcie ČR kakaových bôbov modelom 3. stupňa



Obr. 33: Odhad trendovej funkcie ČR kakaových bôbov modelom 4. stupňa

## C Fisherov test aplikovaný na hodnoty periodogramu ceny kakaových bôbov - hodnoty

	Dĺžka periódy	Fisherová štatistika	Kritická hodnota		Dĺžka periódy	Fisherová štatistika	Kritická hodnota
1	149.00	0.2588	0.062864	10	16.56	0.0818	0.066377
2	74.50	0.2238	0.063235	11	13.55	0.0810	0.066792
3	99.33	0.2180	0.06361	12	27.09	0.0796	0.067214
4	59.60	0.2088	0.06399	13	21.29	0.0791	0.067641
5	42.57	0.1784	0.064375	14	17.53	0.0800	0.068074
6	49.67	0.1253	0.064765	15	10.28	0.0791	0.068513
7	298.00	0.0946	0.06516	16	22.92	0.0818	0.068958
8	29.80	0.0810	0.065561	17	12.96	0.0711	0.069409
9	14.19	0.0791	0.065966	18	9.3125	0.0680	0.069867

Tabuľka 16: Testovanie významnosti periód

**D Hodnoty parametrov  $\delta_j, j = 1, \dots, 4$  a  $\beta_i, i = 1, 2, \dots, 39$  modelu popisujúceho vývoj ceny ka-kaových bôbov**

$\delta_0$	1602.4	$\beta_{19}$	46.052
$\delta_1$	1.5898	$\beta_{20}$	-29.225
$\delta_2$	0.057699	$\beta_{22}$	-51.724
$\beta_1$	214.66	$\beta_{23}$	48.643
$\beta_3$	167.96	$\beta_{25}$	43.309
$\beta_4$	-47.898	$\beta_{26}$	19.463
$\beta_5$	29.708	$\beta_{27}$	31.287
$\beta_6$	151.51	$\beta_{28}$	-33.092
$\beta_7$	-87.826	$\beta_{29}$	41.709
$\beta_8$	-99.164	$\beta_{31}$	24.632
$\beta_9$	22.957	$\beta_{32}$	35.064
$\beta_{10}$	-105.8	$\beta_{33}$	-20.987
$\beta_{11}$	-42.551	$\beta_{34}$	-31.661
$\beta_{12}$	72.257	$\beta_{35}$	28.592
$\beta_{13}$	-86.502	$\beta_{36}$	-21.502
$\beta_{15}$	36.305	$\beta_{37}$	30.576
$\beta_{16}$	47.115	$\beta_{38}$	-15.919
$\beta_{18}$	-54.282	$\beta_{39}$	-34.465

Tabuľka 17: Výsledné hodnoty parametrov  $\delta_i, i = 1, \dots, 4$  a  $\beta_j, j = 1, \dots, 39$

**Poznámka D.1.** Hodnoty parametrov  $\beta_i$ , ktoré v tabuľke nie sú uvedené, sú rovné 0.

## E Priložené CD

### Obsah priloženého CD

- `Skutočné_a_odhadnuté_hodnoty_ČR_ceny_bavlny.pdf`
  - obsahuje tabuľku skutočných a odhadnutých hodnôt ceny bavlny pre všetky skúmané časové okamihy a taktiež hodnotu reziduí pre každý časový okamih
- `Hodnoty_peridogramu_ČR_ceny_bavlny_očisteného_od_trendu.pdf`
  - obsahuje tabuľku s dĺžkami periód a prislúchajúcimi hodnotami periodogramu časového radu ceny bavlny očisteného od trendu
- `Skutočné_a_odhadnuté_hodnoty_ČR_ceny_kakaových_bôbov.pdf`
  - obsahuje tabuľku skutočných a odhadnutých hodnôt ceny kakaových pre všetky skúmané časové okamihy a taktiež hodnotu reziduí pre každý časový okamih
- `Hodnoty_peridogramu_ČR_ceny_kak_bôbov_očisteného_od_trendu.pdf`
  - obsahuje tabuľku s dĺžkami periód a prislúchajúcimi hodnotami periodogramu časového radu ceny kakaových bôbov očisteného od trendu
- `bavlna.m`
  - obsahuje zdrojový kód so všetkými výpočtami použitými pri práci s časovým radom ceny bavlny
- `kakao_boby.m`
  - obsahuje zdrojový kód so všetkými výpočtami použitými pri práci s časovým radom ceny kakaových bôbov
- `Štatistická_analýza_časového_radu_cien_komodít_na_burze.pdf`
  - obsahuje celú prácu v elektronickej podobe

## Literatura

- [1] ANDĚL, J., Statistická analýza časových řad, 1. vydanie. SNTL, Praha, 1976.
- [2] ANDĚL, J., Statistické metody, 3. vydanie. MATFYZPRESS, Praha, 2003.
- [3] ANDĚL, J., Základy matematické statistiky, 1. vydanie. MATFYZPRESS, Praha, 2005.
- [4] BUDÍKOVÁ, M., KRÁLOVÁ, M., MAROŠ, B., Průvodce základními statistickými metodami, 1. vydanie. Grada, Praha, 2010.
- [5] CIPRA, T., Analýza časových řad s aplikacemi v ekonomii, 1. vydanie. SNTL, Praha, 1986.
- [6] COUFALOVÁ, P., Předpovědi v ekonomických časových řadách a jejich aplikace na časovou řadu HDP České republiky. Univerzita Palackého v Olomouci, Přírodovědecká fakulta, Katedra matematické analýzy a aplikací matematiky, 2010, Vedoucí diplomové práce RNDr. Tomáš Fürst, Ph.D.
- [7] DĚDIČ, J., Burza cenných papírů a komoditní burza, Prospektrum, Praha, 1992.
- [8] FONTANILLS, G. A., Getting started in commodities, 1. vydanie. John Wiley & Sons Inc., New Jersey, 2007.
- [9] FORBELSKÁ, M., Stochastické modelování jednorozměrných časových řad, 1. vydanie. Masarykova univerzita, Brno, 2009.
- [10] HINDLS, R., KAŇOKOVÁ, J., NOVÁK, I., Metody statistické analýzy pro ekonomy, 1. vydanie. MANAGEMENT PRESS, Praha, 1997.
- [11] KOLB, R. W., Financial Derivatives, Kolb Publishing Co., Miami, 1993.
- [12] KUBÁČEK, L., KUBÁČKOVÁ, L., Statistika a metrologie, 1.vydanie. Vydavatelství UP, Olomouc, 2000.
- [13] MELOUN, M., MILITKÝ, J., Statistická analýza experimentálních dat, 2. vydanie. Academia, Praha, 2004.
- [14] MONTGOMERY, D. C., PECK, E. A., VINING, G. G., Introduction to linear regression analysis, 3. vydanie. Wiley, Chichester, 2001.
- [15] MURPHY, J. J., Technical analysis of the financial markets, Prentice Hall, New York, 1999.



- [16] URBANÍKOVÁ, M., Využitie finančných derivátov pri riadení vybraných druhov rizík. In Finanční řízení podniků a finančních institucí. Ostrava, VŠB-TU Ostrava, Ekonomická fakulta, 2009, 445 - 457.
- [17] VESELÁ, J., Analýzy trhu cenných papírů II. díl: Fundamentální analýza, 1. vydanie. Oeconomica, Praha, 2003.
- [18] VESELÁ, J., Burzy a burzovní obchody - výchozí texty ke studiu, 1. vydanie. Oeconomica, Praha, 2005.

*Internetové zdroje*

- [19] Webové stránky burzy CME Globex  
[http://www.cmegroup.com/company/files/CME\\_cap\\_brochure\\_printer\\_finals.pdf](http://www.cmegroup.com/company/files/CME_cap_brochure_printer_finals.pdf) [21.3.2011]
- [20] Webové stránky burzy NYSE Euronext Liffe  
<http://www.nyse.com/about/history/1089312755484.html> [1.3.2011]
- [21] Webové stránky burzy ICE - InterContinental Exchange  
<https://www.theice.com/history.jhtml> [1.3.2011]
- [22] Financial Markets Series, Commodities trading, March 2011  
<http://www.thecityuk.com/media/216938/commodities%20trading%202011.pdf> [15.3.2011]
- [23] XTB vzdelávanie - Fundamentálna analýza  
[http://www.xtb.sk/vzdelanie/seminare/celodenne\\_seminare/repository/Downloads/Basic%20seminars%20ppts/XTBSR13%20-%20Fundamentalna%20analyza%20I.pdf](http://www.xtb.sk/vzdelanie/seminare/celodenne_seminare/repository/Downloads/Basic%20seminars%20ppts/XTBSR13%20-%20Fundamentalna%20analyza%20I.pdf) [21.3.2011]
- [24] Článok : Rok 2010: Komoditným lídrom je bavlna, jej cena vzrástla dvojnásobne  
<http://www.investujeme.sk/clanky/rok-2010-komoditnym-lidrom-bavlna-jej-cena-vzrastla-dvojnashobne/> [18.3.2011]
- [25] Článok : Kakao láme rekordy. Môže za to Brit, ktorý kúpil kakao za miliardu dolárov  
<http://ekonomika.sme.sk/c/5472166/kakao-lame-rekordy-moze-za-to-brit-ktory-kupil-kakao-za-miliardu-dolarov.html> [18.3.2011]
- [26] Dáta - ceny bavlny  
<http://www.indexmundi.com/commodities/?commodity=cotton&months=360> [23.3.2010]

- [27] Dáta - ceny kakaových bôbov  
<http://www.indexmundi.com/commodities/?commodity=cocoa-beans&months=360> [23.10.2010]