

MORAVSKÁ VYSOKÁ ŠKOLA OLMOUC

Ústav informatiky a aplikované matematiky

Časové řady a jejich vyhodnocování pomocí softwaru

Mathematica

Time series and their solving by means of software Mathematica

Bakalářská práce

Veronika Šauerová

Vedoucí práce: Mgr. Jan Wossala, Ph.D.

Olomouc 2020

PROHLÁŠENÍ

Prohlašuji, že bakalářskou práci „Časové řady a jejich vyhodnocování pomocí softwaru Mathematica“ jsem vypracovala samostatně a s použitím jen uvedené literatury a zdrojů. Tištěná verze se shoduje s elektronickou verzí na CD nosiči a elektronickou vloženou do IS/STAG.

V Olomouci dne 18. 3. 2020

Veronika Šauerová

PODĚKOVÁNÍ

Tímto děkuji vedoucímu své bakalářské práce panu Mgr. Janu Wossalovi, Ph.D. a paní RNDr. Vratislavě Mošové, CSc. za cenné rady a připomínky pro zpracování daného tématu.

ZADÁNÍ BAKALÁŘSKÉ PRÁCE

(PROJEKTU, UMĚLECKÉHO DÍLA, UMĚLECKÉHO VÝKONU)

Jméno a příjmení: **Veronika Šauerová**
Osobní číslo: **M17048**
Studijní program: **B6208 Ekonomika a management**
Studijní obor: **Podniková ekonomika a management**
Název tématu: **Časové řady a jejich vyhodnocování pomocí softwaru Mathematica**
Téma anglicky: **Time Series and their Solving by means of Software Mathematica**
Zadávací katedra: **Ústav informatiky a aplikované matematiky**

Z á s a d y p r o v y p r a c o v á n í :

1. Úvod - V úvodu bude stanoven cíl práce a představeno téma časových řad.
2. Teoretická část - Teoretická část se bude zabývat časovými řadami a softwarem Mathematica.
3. Metodická část - V této části budou popsány vybrané metody pro vyhodnocování časových řad a postup řešení v softwaru Mathematica.
4. Praktická část - V praktické části budou metody vyhodnocování aplikovány na konkrétní příklady pomocí softwaru Mathematica.
5. Závěr - Shrnutí a zhodnocení práce a dosažení stanovených cílů.

Rozsah grafických prací:

Rozsah pracovní zprávy:

Forma zpracování bakalářské práce: **tištěná/elektronická**

Seznam odborné literatury:

1. ABELL, M. L., BRASELTON J. P.: *Mathematica by example*. Elsevier London 2009, ISBN 978-0-12-374318-3
2. ARTL, J., ARTLOVÁ, M., RUBLÍKOVÁ, E.: *Analýza ekonomických časových řad s příklady*. VŠE Praha, 2004. ISBN:978-80-247-3243-5
3. SIEGEL, A. F.: *Practical Business Statistics*. Elsevier, 2011. ISBN 9780123852083
4. WALKER, I. *Výzkumné metody a statistika*. Praha: Grada, 2013. ISBN 978-80-247-3920-5

Vedoucí bakalářské práce:

Mgr. Jan WOSSALA, Ph.D.

Ústav informatiky a aplikované matematiky

Datum zadání bakalářské práce: **24. května 2019**

Termín odevzdání bakalářské práce: **31. března 2020**

Podpis studenta: *Veronika Šamr* Datum: *6. 9. 2019*

Podpis vedoucího práce: *Jan Wossala* Datum: *28. 6. 2019*

Kováčičinová
Mgr. Irena KOVAČIČINOVÁ
prorektorka



Mgr. Veronika ŘÍHOVÁ, Ph.D.
manažer ústavu

Říhová

V Olomouci dne 19. června 2019

Obsah

ÚVOD	8
I. TEORETICKÁ ČÁST	11
1 SOFTWARE MATHEMATICA	11
1.1 SOFTWARE MATHEMATICA A JEHO STRUKTURA	11
1.2 PRÁCE V PROSTŘEDÍ NOTEBOOKU SOFTWARE MATHEMATICA	12
2 ČASOVÉ ŘADY	15
2.1 VKLÁDÁNÍ, GENEROVÁNÍ, PŘEVOD A EXPORTOVÁNÍ DAT	15
2.2 GRAFY, CHARAKTERISTIKY A MÍRY DYNAMIKY	16
2.3 ZÁKLADNÍ CHARAKTERISTIKY ČASOVÝCH ŘAD	20
2.4 METODY ANALÝZY ČASOVÝCH ŘAD	22
3 MODELOVÁNÍ ČASOVÝCH ŘAD METODOU DEKOMPOZICE	24
3.1 MODELOVÁNÍ TRENDU	25
3.2 TRENDOVÉ FUNKCE A JEJICH VÝBĚR	26
3.3 INTERPOLAČNÍ KRITÉRIA	28
3.4 KLOUZAVÉ PRŮMĚRY	33
3.5 PŘEDPOVĚDI V ČASOVÝCH ŘADÁCH	34
4 ANALÝZA ČASOVÝCH ŘAD NA ZÁKLADĚ BOXOVY-JENKINSOVY METODOLOGIE	36
4.1 ZÁKLADNÍ PRINCIPY	36
4.2 MODEL Y A JEJICH IDENTIFIKACE	37
4.3 PROGNOZOVÁNÍ POMOCÍ MODELU ARIMA	42
II. PRAKTICKÁ ČÁST	43
5 ZÍSKÁVÁNÍ DAT PRO PRÁCI V PROGRAMU MATHEMATICA	44
PŘÍKLAD 5.1 (GENEROVÁNÍ VLASTNÍCH DAT)	44
PŘÍKLAD 5.2 (IMPORT DAT ZE SOUBORU)	46
6 ZÁKLADNÍ CHARAKTERISTIKY ČASOVÝCH ŘAD	48
PŘÍKLAD 6.1 (PROSTÝ ARITMETICKÝ PRŮMĚR INTERVALOVÉ ČASOVÉ ŘADY)	48
PŘÍKLAD 6.2 (CHRONOLOGICKÝ ARITMETICKÝ PRŮMĚR OKAMŽIKOVÉ ČASOVÉ ŘADY)	49
PŘÍKLAD 6.3 (MÍRY DYNAMIKY)	51
7 METODA DEKOMPOZICE	54
PŘÍKLAD 7.1 (MODELOVÁNÍ TRENDŮ A JEJICH VZÁJEMNÉ POROVNÁNÍ)	54
8 KLOUZAVÉ PRŮMĚRY	61
PŘÍKLAD 8.1 (OČIŠTĚNÍ ŘADY OD SEZÓNŇNÍCH VLVIVŮ)	61
9 BOX-JENKINSOVA METODOLOGIE	66
PŘÍKLAD 9.1 (KONSTRUKCE ARIMA MODELU)	66
PŘÍKLAD 9.2 (PROGNOZOVÁNÍ ČASOVÉ ŘADY)	70

ZÁVĚR	73
LITERATURA A ZDROJE	75
SEZNAM ZKRATEK	77
SEZNAM ZNAČEK	78
SEZNAM TABULEK	79
SEZNAM OBRÁZKŮ	80
SEZNAM PŘÍLOH	82
PŘÍLOHY	83
ANOTACE	87

Úvod

Předmětem této bakalářské práce je soubor příkazů a formulí softwaru Mathematica využitelných při analýze časových řad. Tyto příkazy a formule budou aplikovány na teorii časových řad a budou demonstrovány na konkrétních příkladech. Práce bude zaměřena na základní popisné charakteristiky, pomocí nichž lze získávat informace o vlastnostech časové řady jako jednoho celku. Dále budou hledány a testovány modely s využitím vybraných metod. Budou také provedeny odhady budoucího vývoje některých časových řad.

Časovou řadou lze nazvat každou chronologicky uspořádanou posloupnost pozorovaných dat. Analýza časových řad má široké využití v různých oborech. Pozorovaná data se zkoumají např. v ekonomii, biologii, lékařství, meteorologii či technice. Mohou se týkat i vývoje cen akcií na burze, denních teplot, počtu jedinců určitého druhu živočichů atd. Časové řady mohou být také cenným nástrojem pro řešení rozhodovacích úloh.

Analýza časových řad přináší informace o vývoji zkoumaných hodnot. Popisuje jejich charakteristické vlastnosti, umožňuje získat chybějící data či odhadovat budoucí hodnoty řady na základě poznatků z minulosti. Pomáhá určit nezávislost časových řad, popřípadě příčiny jevů ovlivňujících jejich vývoj. Hlavním úkolem analýzy je pak konstrukce modelu vystihujícího průběh časové řady, který umožní získ požadovaných informací.

Stávající práce bude zaměřena na dvě základní metody používané k analýze časové řady a k nalezení matematického modelu, který danou řadu vystihuje co nejpřesněji. Jedná se o metodu dekompozice a metodu založenou na Box-Jenkinsově metodologii. Obě metody budou popsány dále.

Výpočty bývají často velmi zdlouhavé a složité, proto se přistupuje k využití IT techniky, která je provádí rychleji a také s větší přesností. Byla vyvinuta řada počítačových programů, pomocí nichž lze časové řady zkoumat. K jednodušším výpočtům lze využít Microsoft Excel. Existují ale také softwary jako STATISTICA, SPSS, S+, STATGRAPHICS zaměřené speciálně na statistické výpočty a skupina programů, které slouží k širší škále vědecko-technických výpočtů. K posledním jmenovaným patří i software Mathematica, na který bude práce zaměřena.

Software Mathematica jsem pro tuto práci vybrala především díky dřívějším zkušenostem s jeho používáním a také pro jeho širokou škálu výpočtových možností, které nabízí nejen pro statistiku, ale i pro další matematické obory. Vlastní výpočet v programu se uskutečňuje pomocí předdefinovaných příkazů, ty však nepokrývají všechny možnosti zkoumání. Proto je nutné v některých případech tvořit vlastní příkazové formule.

Jako ukázka použití příkazů bylo pro tuto bakalářskou práci vybráno několik časových řad z různých odvětví. Mezi vybrané řady patří: Výdaje vztahující se k výzkumu a vývoji na území České republiky v letech 1989-2018, Počet subjektů provádějící výzkum a vývoj v České republice a také Celkový počet hostů hromadných ubytovacích zařízení v České republice v letech 2012-2018.

Cílem stávající bakalářské práce je zjistit, které příkazy softwaru Mathematica jsou využitelné pro vyhodnocování časových řad, popsat je a demonstrovat na příkladech s konkrétními časovými řadami. U zvolených řad budou zkoumány jejich vlastnosti na základě výpočtu charakteristik a měř dynamiky. Dále pak pro ně budou vyhledávány a konstruovány vhodné modely. Ty budou získávány pomocí metody dekompozice a Box-Jenkinsovy metodologie.

Práce obsahuje řadu faktů týkajících se časových řad a softwaru Mathematica, proto by mohla posloužit všem, kteří se časovými řadami zabývají, i uživatelům softwaru Mathematica. Uvedené příklady mohou být použity jako návod pro práci s časovými řadami a jejich analýzu v tomto programu. V práci je možné zjistit, jakým způsobem lze v softwaru Mathematica časové řady vytvářet, vkládat je a exportovat. Příklady také ukazují, jak lze očistit časovou řadu od sezónních vlivů, získat její základní charakteristiky, modelovat a ověřovat trendovou funkci, konstruovat ARIMA model, předpovídat budoucí hodnoty atd. Přínos práce by mohl spočívat také v interpretaci konkrétních výsledků analýzy časových řad. Získají se tak informace o vývoji hodnot výdajů na výzkum a vývoj, o počtu subjektů, které takový výzkum provádějí nebo o počtu hostů hromadných ubytovacích zařízení v České republice. Dle Českého statistického úřadu výzkum a vývoj je významným prvkem zvyšování produktivity, ekonomického růstu a zaměstnanosti.¹ Proto by predikce vývoje výdajů na výzkum a vývoj mohly být cenným poznatkem pro řešení různých rozhodovacích úloh.

Text práce bude dále rozčleněn do dvou základních částí. První z nich bude teoretická, která bude začínat stručným úvodem do softwaru Mathematica. Zaměřena však bude především na matematický rozbor charakteristik časových řad a vybrané metody pro konstrukce modelů časových řad. Pod každou kapitolou či podkapitolou budou uvedeny příslušné příkazy používané k analýze v programu software Mathematica. Od textu týkajícího se časových řad budou odděleny označením „SM – ukázka syntaxe příkazů“, kde SM znamená software Mathematica. Ve druhé – praktické – části bude pro řešení praktických úloh využívána teorie

¹ Srov. Wolfram Research - Company Background. *Wolfram: Computation Meets Knowledge* [online]. Wolfram, 2020 [cit. 2020-03-14]. Dostupné z: <https://www.wolfram.com/company/background.html>.

z teoretické části práce. Popsané příkazy budou používány pro zkoumání vlastností a vývoje vybraných časových řad. Jednotlivé příklady budou zařazeny do odstavců s názvy těch podkapitol z teoretické části práce, ke kterým se řešená problematika vztahuje.

Stěžejní metodikou bude analýza a modelování, které jsou přímo spojené se samotným cílem této práce. Analýza časových řad bude sloužit k získávání modelů a na jejich základě bude možné predikovat další vývoj ukazatelů. Dále bude použita demonstrace při ukázkách na příkladech a komparace pro porovnávání skutečných a odhadnutých hodnot modelů.

I. Teoretická část

1 Software Mathematica

Wolfram Mathematica je počítačový program, který se využívá jako nástroj k provádění výpočtů a vizualizaci dat. Jeho první verze vznikla v roce 1988 zásluhou Stephena Wolframa a skupiny jeho spolupracovníků. Od té doby prošel řadou změn. Nyní je vedle dalších produktů poskytován firmou Wolfram Research. Má široké využití při řešení různých typů úloh. Nejprve sloužil při řešení matematických, fyzikálních a inženýrských úloh, později jej začaly aplikovat v dalších technických, ale i jiných oborech.

1.1 Software Mathematica a jeho struktura

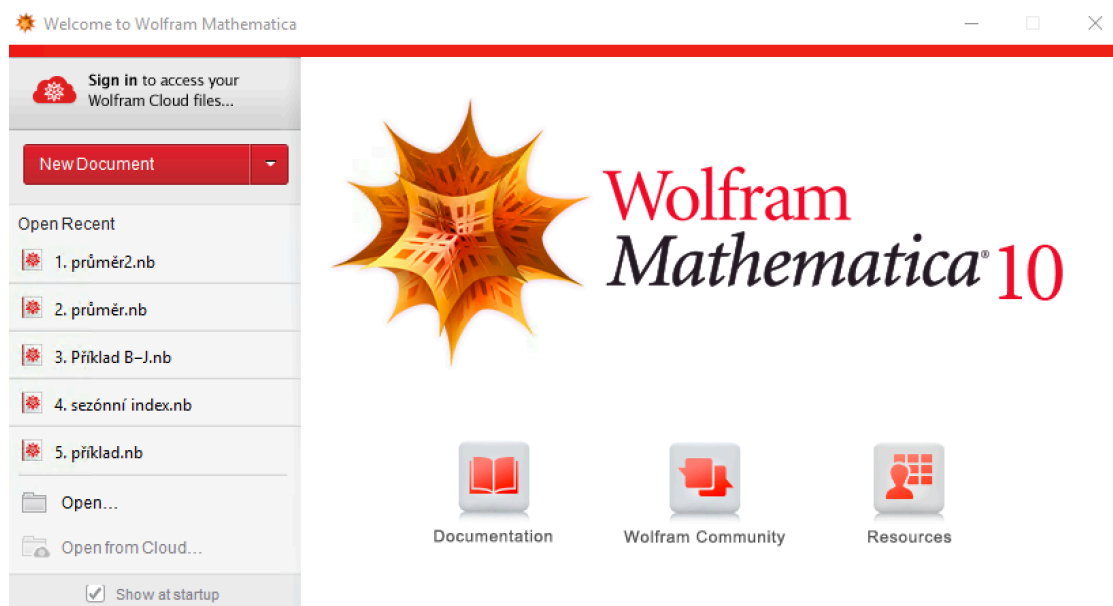
Software Mathematica patří mezi CAS tj. počítačové algebraické systémy. Jedná se o systémy, které jsou schopné počítačově zpracovávat matematické symboly. Software Mathematica je podporován operačními systémy Windows Microsoft, Apple Mac OS i Linuxem. Přestože celý program je v angličtině, je menu dostupné i v jiných jazycích, což ovšem podléhá licenčním podmínkám. Dále tento program používá programovací jazyk Wolfram.²

Program pracuje s rovnicemi a jejich soustavami, vektory, maticemi, reálnými čísly, komplexními čísly atd. Lze jej použít jako kalkulačku i jako nástroj pro řešení složitých a numericky náročných úloh. Vytvořené výstupy mohou mít písemnou, grafickou, zvukovou i animační podobu. Program disponuje interaktivní nápovědou, kterou lze využít také v režimu offline. Poskytuje vysvětlení jednotlivých příkazů a ukazuje příklady jejich použití.

Po otevření programu se zobrazí okno, ve kterém lze nalézt novinky ve vývoji programu i seznam posledních otevřených souborů. V okně si lze také vybrat typ nového souboru např. notebook, prezentace atd. Aplikace je členěna na hlavní nabídku a další okna, se kterými lze manipulovat zvlášť (*Obrázek 1*).³

² Srov. Wolfram Research - Company Background. *Wolfram: Computation Meets Knowledge* [online]. Wolfram, 2020 [cit. 2020-03-14]. Dostupné z: <https://www.wolfram.com/company/background.html>.

³ Srov. BOUŠKA, Martin. *MATHEMATICA - příručka s příklady pro učitele a studenty*, s. 3.



Obrázek 1 Vstupní okno do programu

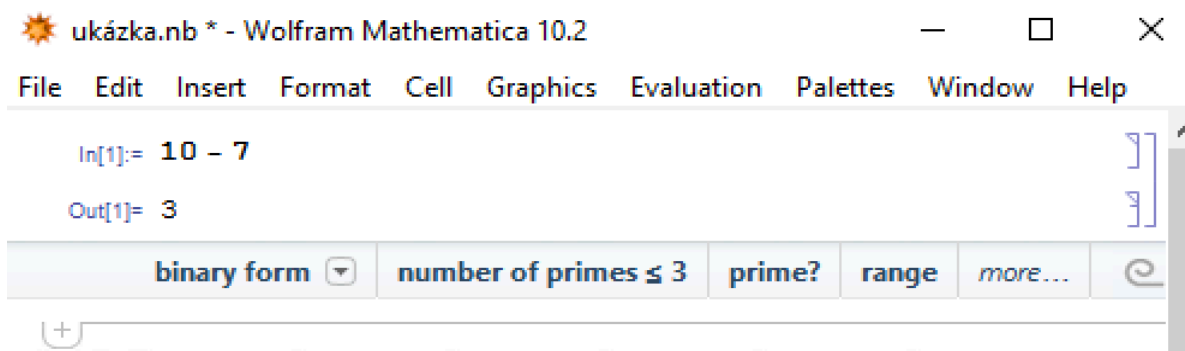
Zdroj: Vlastní zpracování

Program Mathematica se skládá ze tří základních částí. První z nich se nazývá MathKernel a plní funkci výpočtového jádra. Principem jádra je provádění přesných výpočtů. Další částí je Front End, který umožňuje uživateli komunikovat s jádrem. Slouží ke vpisování příkazů, které následně odesílá do jádra a zobrazí vyhodnocené výsledky. Třetí a poslední část je Package, která obsahuje méně často používaná rozšíření. Celý program je totiž velmi rozsáhlý a běžný uživatel jej nemusí mít stále k dispozici. Aby rozšíření nezatěžovala aplikaci, jsou uložena v balíčku.⁴

1.2 Práce v prostředí notebooku softwaru Mathematica

Práce v programu se dělí na buňky, ty mohou být textové nebo výpočetní. Výpočetní buňky se dále dělí na vstupní a výstupní (Obrázek 2). Do vstupních buněk se vkládají příkazy a jsou označeny `In[]`. Po stisknutí `Shift+Enter` se pak výstupní buňky vyhodnocují a značí se `Out[]`. S buňkami je možné provádět různé operace, např. slučovat, rozdělovat, určit, aby nebyly vyhodnoceny atd. Označení buněk je zobrazeno na pravé straně dokumentu.

⁴ Srov. BOUŠKA, Martin. *MATHEMATICA - příručka s příklady pro učitele a studenty*, s. 3.



Obrázek 2 Ukázka výpočtu

Zdroj: Vlastní zpracování

Do vstupních buněk je možné zadávat matematické operátory i různé příkazy. Mathematica rozlišuje velká a malá písmena. To je důležité právě při zadávání příkazů, kde první písmeno každého slova musí být velké a ostatní jsou malá. Parametry příkazů se vkládají do hranatých závorek např. $RandomReal[x_{max}]$. Složené závorky pak slouží k vymezení seznamů a kulaté k uzávkování matematických výrazů. Textové popisky např. `os` je nutné vkládat do uvozovek. Pro případ, kdy není žádoucí, aby se zobrazila výstupní buňka, se používá středník, který se připojí na konec příslušné vstupní buňky. Jména vytvořených výstupů s sebou nesou informace, které umožňují se na příslušné výstupy odvolávat. Např. není nutné znovu vypisovat již dříve vytvořený a pojmenovaný seznam hodnot, ale stačí připsat jeho jméno.⁵

SM – ukázka syntaxe příkazů

- Práce se seznamy:

`Partition[list,n,d]`

rozdělí seznam *list* do menších podseznamů o počtu členů *n* s posunem o *d* členů

`Part[list,i]` nebo `list[[i]]`

vybere *i*-tý člen ze seznamu *list*

`Take[list{i,j}]`

vybere ze seznamu *list* *i*-tý až *j*-tý člen

⁵ Srov. Wolfram Language & System Documentation Center. *Wolfram: Computation Meets Knowledge* [online]. Wolfram, 2020 [cit. 2020-03-14]. Dostupné z: <https://reference.wolfram.com/language/>.

Transpose[list]

transponuje první dvě úrovně seznamu *list*

Join[list₁,list₂,...]

spojí seznamy *list₁,list₂,...* v jeden seznam

Length[list]

určí počet členů seznamu *list*

- Zobrazení:

TableForm[list,options]

tvoří tabulku ze seznamu *list* s možností dalších voleb *options*

Show[graphics,options]

vykreslí grafické objekty *graphics* s možností dalších voleb *options*

Grid[list,options]

tvoří mřížku k seznamu *list* s možností dalších voleb *options*

- Další:

N[expr,n]

vyčíslí výraz *expr* na *n* desetinných míst

2 Časové řady

Ze všeho nejdříve je potřeba si vymežit pojem časové řady. Časová řada vzniká pozorováním určitého ukazatele v čase. Lze proto říci, že se jedná o posloupnost dat chronologicky uspořádanou.⁶ Aby bylo možné řadu analyzovat, je nutné dodržet pořadí členů. V opačném případě by docházelo ke ztrátám důležitých informací.

Na základě analýzy časových řad lze předpovídat další vývoj vybraných ukazatelů nebo odhalovat příčiny vzniklých jevů. Nejprve je ale nutné řady přesně specifikovat. Např. lze uvažovat řady, které sledují vývoj inflace v České republice mezi lety 2002-2018, počet obyvatel v okrese Kladno mezi lety 2000 a 2018, průměrný věk obyvatel na území České republiky od roku 1990 do roku 2019, minimální denní teploty v okrese Olomouc během roku 2018, celkové rodinné měsíční výdaje během roku 2019 atd. Jak je vidět, údaji závislémi na čase mohou být ukazatele ekonomické, demografické i meteorologické, mohou se ale také týkat domácího prostředí.⁷

2.1 Vkládání, generování, převod a exportování dat

V softwaru Mathematica existuje několik možností, jak získat časovou řadu. Program ji umí vygenerovat sám pomocí příkazů *RandomReal* nebo *RandomInteger*. Tyto příkazy se liší podle množin, ze kterých jsou náhodná čísla vybírána. První příkaz, *RandomReal*, vybírá hodnoty z množiny reálných čísel a *RandomInteger* z množiny celých čísel. Těmto hodnotám pak stačí přiřadit časovou složku příkazem *TimeSeries*.

Další možností, jak řadu získat, je přenést ji do programu z jiného souboru nebo webové stránky pomocí příkazu *Import*. Naopak, pokud je potřeba data exportovat, lze použít příkaz *Export*. Jak vkládání tak exportování dat lze provádět ve formě seznamu, tabulky atd. Program podporuje pro soubory, ze kterých a do kterých je import i export prováděn, širokou škálu formátů.⁸ Tyto formáty se zobrazí po vyhodnocení buňky s textem „*\$ImportFormats*“ nebo „*\$ExportFormats*“.⁹

⁶ Srov. CIPRA, Tomáš. *Analýza časových řad s aplikacemi v ekonomii*, s. 9.

⁷ Srov. LITSCHMANNOVÁ, Martina. *Úvod do analýzy časových řad*, s. 2.

⁸ WELLIN, Paul a Harry CALKINS. *M101 A First Course in Mathematica.*, s. 266.

⁹ Srov. Wolfram Language & System Documentation Center. *Wolfram: Computation Meets Knowledge* [online]. Wolfram, 2020 [cit. 2020-03-14]. Dostupné z: <https://reference.wolfram.com/language/>.

SM – ukázka syntaxe příkazů

RandomReal[range,n]

generuje n reálných čísel v zadaném rozmezí *range*

RandomInteger[range,n]

generuje n celých čísel v zadaném rozmezí *range*

Range[imin,imax,d]

generuje posloupnost čísel od i_{min} po i_{max} s diferencí d

TimeSeries[values,{times}]

určí časovou řadu přiřazením časové složky *times* hodnotám *values*

TemporalData[{v1,v2,...},times]

přiřadí hodnoty ze seznamů v_1, v_2, \dots časové složce *times*

Import["file","List"]

importuje do programu hodnoty ze souboru *file* ve formě seznamu *List*

Export["file",list,"Table"]

exportuje seznam hodnot *list* do souboru *file* ve formě tabulky *Table*

2.2 Grafy, charakteristiky a míry dynamiky

Časová řada je chronologicky uspořádaná posloupnost ukazatelů. Nadále bude pro hodnoty časové řady používáno značení y_t , kde $t = 1, 2, \dots, T$.

Na rozlišení časových řad lze pohlížet z různých úhlů např. podle charakteru dat lze řady rozdělit na časové řady intervalové, okamžikové a řady odvozených charakteristik. Toto dělení je důležité při způsobu počítání aritmetických průměrů, které se u jednotlivých typů řad liší.

Příkladem intervalové časové řady může být výše nákladů dané firmy v jednotlivých měsících ve vybraném roce. Zde se vyhodnocuje celková výše nákladů za měsíc. Pro výpočet aritmetického průměru se používá tzv. prostý aritmetický průměr:

$$\bar{y} = \frac{y_1 + y_2 + \dots + y_T}{T}.$$

V softwaru Mathematica lze pro výpočet použít příkaz *Mean*. Je třeba ještě připomenout, že pokud se délka intervalů liší, je zapotřebí data přepočítat na srovnatelné hodnoty. Proto se provádí tzv. očištění časové řady od kalendářních vlivů.¹⁰

Okamžikové časové řady se vztahují k určitému okamžiku a data nemá smysl sčítat. Příkladem takové řady může být počet zaměstnanců firmy ke každému prvnímu dni v měsíci ve vybraném roce. Pro výpočet průměru okamžikové časové řady se používá chronologický aritmetický průměr:

$$\bar{y} = \frac{\frac{y_1 + y_2}{2} + \frac{y_2 + y_3}{2} + \dots + \frac{y_{T-1} + y_T}{2}}{T - 1}.$$

Software Mathematica nenabízí pro výpočet samostatný příkaz, je tedy nutné vytvořit formuli složením z více příkazů. To bude učiněno v praktické části práce. Chronologický aritmetický průměr je třeba upravit, pokud je délka mezi jednotlivými časovými okamžiky různá. V případě délek d_1, \dots, d_{T-1} se aplikuje vážený chronologický aritmetický průměr:

$$\bar{y} = \frac{\frac{y_1 + y_2}{2}d_1 + \frac{y_2 + y_3}{2}d_2 + \dots + \frac{y_{T-1} + y_T}{2}d_{T-1}}{d_1 + d_2 + \dots + d_{T-1}}. \quad 11$$

Časové řady odvozených charakteristik se nejčastěji získávají součtem nebo dělením předchozích typů řad. Jedná se např. o řady hrubého domácího produktu na jednoho obyvatele apod.¹²

Ekonomické časové řady lze rozlišovat také na dlouhodobé a krátkodobé. Dlouhodobé řady bývají sledovány v ročních intervalech a krátkodobé řady v intervalech kratších než jeden rok.¹³

¹⁰ Srov. ARLT, Josef, Markéta ARLTOVÁ a Eva RUBLÍKOVÁ. *Analýza ekonomických časových řad s příklady*, s. 13.

¹¹ Srov. LITSCHMANNOVÁ, Martina. *Úvod do analýzy časových řad*, s. 3.

¹² Srov. ARLT, Josef, Markéta ARLTOVÁ a Eva RUBLÍKOVÁ. *Analýza ekonomických časových řad s příklady*, s. 7.

¹³ Srov. ARLT, Josef a Markéta ARLTOVÁ. *Finanční časové řady*, s. 12.

SM – ukázka syntaxe příkazů

Mean[list]

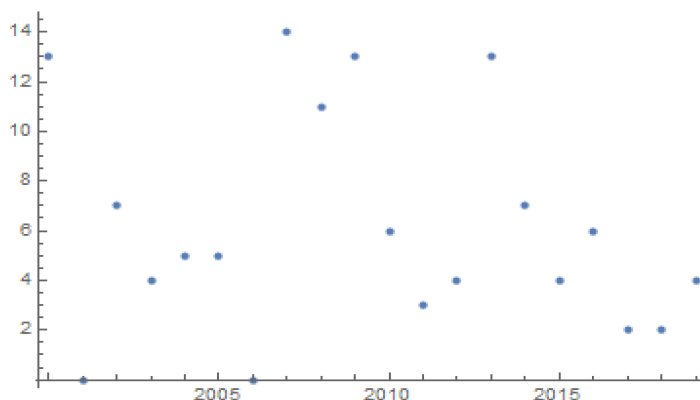
počítá aritmetický průměr ze členů seznamu *list*

Sum[f,{i,l_{min},l_{max}}]

sečte hodnoty funkce *f* o proměnné *i* od *l_{min}* po *l_{max}*

Pro lepší názornost lze časové řady vyjádřit graficky. Existuje několik způsobů, jak pomocí grafu vyjádřit průběh časové řady. Varianta se vybírá podle toho, která charakteristika má být vyzdvížena. Graf je možné tvořit přímo z původních hodnot a u intervalových řad se někdy zobrazují kumulované hodnoty. Nejčastěji se pro znázornění používá bodový (Obrázek 3) nebo spojnicový (Obrázek 4) graf původních hodnot. Pokud mají být ale z grafu patrné jiné informace jako třeba aritmetický průměr, medián, maximum, minimum nebo další, je vhodnější sestojit krabičkový (Obrázek 5) graf. Další možností znázornění je graf sezónních hodnot, který vypovídá o odchylkách skutečných hodnot od aritmetického průměru.¹⁴

Bodový graf lze získat po vyhodnocení příkazu *ListPlot*, spojnicový po vyhodnocení příkazu *ListLinePlot* a krabičkový graf se vykreslí při užití příkazu *BoxWhiskerChart*. Program umožňuje i další úpravy grafu, jako popisování os, vložení legendy, nadpis grafu atd.¹⁵ Příklady uvedených grafů budou ukázány na následujících obrázcích (Obrázek 3-5).



Obrázek 3 Bodový graf

Zdroj: vlastní zpracování

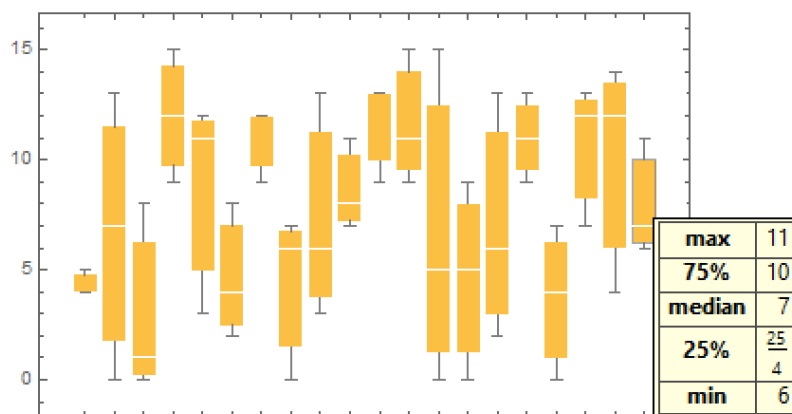
¹⁴ Srov. ARLT, Josef, Markéta ARLTOVÁ a Eva RUBLÍKOVÁ. *Analýza ekonomických časových řad s příklady*, s. 7-12.

¹⁵ Srov. Wolfram Language & System Documentation Center. *Wolfram: Computation Meets Knowledge* [online]. Wolfram, 2020 [cit. 2020-03-14]. Dostupné z: <https://reference.wolfram.com/language/>.



Obrázek 4 Spojnicový graf

Zdroj: vlastní zpracování



Obrázek 5 Krabičkový graf

Zdroj: vlastní zpracování

Jak lze vidět na obrázcích, mezi prvními dvěma grafy není příliš velký rozdíl, třetí z nich se ale výrazně odlišuje. V krabičkovém grafu není možné vidět průběh řady a odhadnout z něj trend, ale je zde možné najít jiné informace, které z bodového a spojnicového grafu nelze bez výpočtu zjistit.

SM – ukázka syntaxe příkazů

ListPlot[{y₁,y₂,...}]

vykreslí bodový graf hodnot y₁,y₂,...

`DiscretePlot[list, {i, imin, imax}`

generuje graf hodnot seznamu *list*, přičemž *i* probíhá od *i_{min}* do *i_{max}*

`ListLinePlot[{y1, y2, ...}]`

proloží body *y₁, y₂, ...* křivkou

`BoxWhiskerChart[{y1, y2, ...}]`

vykreslí krabičkový graf hodnot *y₁, y₂, ...*

`DateListPlot[tserie]`

vykreslí graf časové řady *tserie*

2.3 Základní charakteristiky časových řad

Základní informace o časových řadách lze získat prostřednictvím základních charakteristik. Ty se dělí na popisné charakteristiky a míry dynamiky. K popisným ukazatelům patří již výše zmíněný aritmetický, chronologický a vážený chronologický průměr.

Míry dynamiky charakterizují chování časových řad, z nichž některé budou uvedeny v této podkapitole. Výpočet měr dynamiky vyžaduje splnění následujících kritérií: Intervalové časové řady musí vyhovovat předpokladu stále stejné délky intervalů. Obdobně okamžikové řady podmiňuje kritérium konstantní vzdálenosti mezi okamžiky.¹⁶ Software Mathematica nabízí jednoduchý příkaz jen pro níže uvedenou první míru dynamiky - absolutní přírůstek, pro ostatní budou složené příkazy a formule vytvořeny až v praktické části.

- Absolutní přírůstek, neboli první diference, vyjadřuje rozdíl mezi sousedními dvěma hodnotami:

$$\Delta y_t = y_t - y_{t-1}, \text{ kde } t = 2, 3, \dots, T.^{17}$$

Tato míra dynamiky určuje, o kolik jednotek se změnila hodnota daného ukazatele oproti minulému období. V programu Mathematica je možné ji spočítat pomocí příkazu *Differences*.¹⁸

¹⁶ Srov. LITSCHMANNOVÁ, Martina. *Úvod do analýzy časových řad*, s. 4.

¹⁷ Srov. HINDLS, Richard. *Statistika pro ekonomy*, s. 252.

¹⁸ Srov. Wolfram Language & System Documentation Center. *Wolfram: Computation Meets Knowledge* [online]. Wolfram, 2020 [cit. 2020-03-14]. Dostupné z: <https://reference.wolfram.com/language/>.

- Průměrný absolutní přírůstek je aritmetickým průměrem všech absolutních přírůstků:

$$\bar{\Delta} = \frac{y_T - y_1}{T - 1}.^{19}$$

Vyjadřuje, o kolik jednotek průměrně rostla nebo klesala hodnota ukazatele za jednotku času. K výpočtu lze použít příkazy *Mean* a *Differences*.²⁰

- Relativní přírůstek udává relativní nárůst nebo pokles hodnoty mezi dvěma obdobími.

$$\delta_t = \frac{y_t - y_{t-1}}{y_{t-1}}, \text{ kde } t = 2, 3, \dots, T.^{21}$$

Po vynásobení jedním stem je dán v procentech. Relativní přírůstek pak odpovídá na otázku, o kolik procent se změnila hodnota ukazatele mezi dvěma obdobími. Vypočítat jej lze použitím příkazů *Divide* a *Partition a Times*.²²

- Koeficient růstu:

$$k_t = \frac{y_t}{y_{t-1}}, \text{ kde } t = 2, 3, \dots, T.^{23}$$

Po vynásobení stem vyjadřuje, na kolik procent vzrostla nebo se snížila původní hodnota. Koeficient růstu se v programu Mathematica počítá pomocí stejných příkazů, jak je tomu u výpočtu relativních přírůstků. Pro jeho výpočet slouží stejné příkazy jako pro získání hodnot relativních přírůstků.

¹⁹ Srov. LITSCHMANNOVÁ, Martina. *Úvod do analýzy časových řad*, s. 4.

²⁰ Srov. Wolfram Language & System Documentation Center. *Wolfram: Computation Meets Knowledge* [online]. Wolfram, 2020 [cit. 2020-03-14]. Dostupné z: <https://reference.wolfram.com/language/>.

²¹ Srov. HANČLOVÁ, Jana a Lubor TVRDÝ. *Úvod do analýzy časových řad*, s. 6.

²² Srov. Wolfram Language & System Documentation Center. *Wolfram: Computation Meets Knowledge* [online]. Wolfram, 2020 [cit. 2020-03-14]. Dostupné z: <https://reference.wolfram.com/language/>.

²³ Srov. LITSCHMANNOVÁ, Martina. *Úvod do analýzy časových řad*, s. 4.

- Průměrný koeficient růstu – počítá se pomocí vzorce:

$$\bar{k} = {}^{T-1}\sqrt{k_1 \cdot k_2 \cdot \dots \cdot k_T} = {}^{T-1}\sqrt{\frac{y_T}{y_1}}.^{24}$$

Stejně jako u předchozích dvou měř i průměrný koeficient růstu nemá v programu přímý příkaz a je nutné vytvořit formuli. Ta se může skládat z příkazů *Product* a matematických operátorů.²⁵

Kromě těchto ukazatelů lze časové řady charakterizovat pomocí dalších statistických veličin jako jsou rozptyl nebo kovariance.

SM – ukázka syntaxe příkazů

Differences[list,n]

počítá *n*-té diference seznamu *list*

Mean[list]

viz 2.2

Apply[f,expr,{1}] nebo f @@@ expr

aplikuje funkci *f* na výraz *expr* na úrovni *l*

Divide[x,y]

provede dělení *x/y*

Times[x,y]

provede součin *x·y*

Product[f,{i,i_{min},i_{max}}

vynásobí hodnoty funkce *f* o proměnné *i* od *i_{min}* po *i_{max}*

2.4 Metody analýzy časových řad

Analyzovat průběh časových řad je možné hned několika způsoby. Metoda je vybírána podle několika aspektů. Důležitou roli hraje důvod analýzy, který se pojí s tím, jak, kde

²⁴ Srov. LITSCHMANNOVÁ, Martina. *Úvod do analýzy časových řad*, s. 4.

²⁵ Srov. Wolfram Language & System Documentation Center. *Wolfram: Computation Meets Knowledge* [online]. Wolfram, 2020 [cit. 2020-03-14]. Dostupné z: <https://reference.wolfram.com/language/>.

a k čemu mají být výsledky dále využity. Většinou je analýza zacílená na hledání zákonitostí vývoje časové řady a následnou tvorbu předpovědí. Rozhodují také další hlediska, jako je typ časové řady, zkušenosti analytika nebo podmínky, ve kterých má zkoumání probíhat.²⁶

Tento text se bude zabývat pouze dvěma základními způsoby vytváření modelů časových řad, a to metodou dekompozice a metodou založenou na Box-Jenkinsově metodologii. Metoda dekompozice časových řad spočívá v rozložení časové řady na složky. Dále se pracuje s těmi, které jsou systematické. Jejich analýza využívá při konstrukci grafu principů regresní analýzy. Cíl metody spočívá v identifikaci pravidelností ve vývoji hodnot jednotlivých složek. Vychází z předpokladu, že pozorování na sobě nejsou závislá.

Na rozdíl od předchozí metody, Box-Jenkinsova metodologie je založena na zkoumání nesystematické složky. Vytvářené modely jsou také mnohem adaptabilnější na změny hodnot a pracují i s řadami, které obsahují závislé neboli korelované náhodné veličiny. Tyto náhodné veličiny jsou hlavním předmětem zkoumání. Dalším rozdílem je, že pomocí tohoto přístupu lze modelovat řady, které mají výraznou trendovou a sezónní složku.

²⁶ Srov. KŘIVÝ, Ivan. *Analýza časových řad*, s. 13.

3 Modelování časových řad metodou dekompozice

Metoda dekompozice předpokládá, že je možné časovou řadu rozložit na čtyři složky – trendovou, sezónní, cyklickou a náhodnou. První tři z nich jsou tzv. systematické složky a poslední je tzv. nesystematická složka. Jedním z důvodů, proč se rozklad provádí je, že pak lze časovou řadu očistit od sezónních vlivů, což umožňuje následné srovnávání dvou řad. Lze ji ale také očistit od trendové složky. Pak je totiž lépe vidět, jak velký vliv má na průběh časové řady sezónní složka. Dalším důvodem je fakt, že po realizaci rozkladu je možné lépe předpovídat další vývoj nebo určovat vzniklé zákonitosti.

Trendová složka T_t vypovídá o dlouhodobém vývoji celé řady. Trend může růst, klesat nebo se v řadě nemusí vůbec vyskytovat. Předmětem zkoumání nejsou krátkodobé výkyvy, ale jen dlouhodobé tendence. Vzniká působením sil, které se vyvíjí stejným směrem.

Cyklická složka C_t se pohybuje okolo trendu. Průběh se nepravidelně mění a složka střídavě roste a klesá. Jednotlivé cykly se od sebe liší a mají různou délku, ovšem vždy větší než jeden rok. Další faktor, ve kterém se cykly v průběhu časové řady liší, je amplituda. Vývoj cyklické složky bývá ovlivněn jak ekonomickými (např. hospodářský cyklus) tak neekonomickými faktory (např. změna klimatu). V posledních letech má na vývoj cyklické složky velký vliv technologický pokrok a inovace, ale přesná příčina se zjišťuje obtížně.

Sezónní složka S_t taktéž kolísá kolem trendu, ale je pravidelnější. Výkyvy se opakují ve stejných obdobích v rámci jednoho roku jako např. tržby lyžařských areálů v zimních měsících nebo vyšší teploty v letních měsících. Sezónní složka tedy ovlivňuje časové řady na základě ročních období (jaro, léto podzim, zima) nebo období svátků (Vánoce a další události).

Náhodná neboli reziduální složka a_t vyjadřuje nesystematické odchylky nebo také chyby v měření. Jedná se o hodnoty, které zbydou po odstranění předchozích tří složek. Reziduální složka má často vlastnosti bílého šumu. Bílý šum se skládá z hodnot nezávislých náhodných veličin, jejichž rozptyl je konstantní a střední hodnota je rovna nule.

Časovou řadu lze rozložit dvojím způsobem. Rozklad může mít formu aditivní nebo multiplikativní.

- Aditivní rozklad spočívá v prostém sečtení všech čtyř složek:

$$y_t = T_t + C_t + S_t + a_t, \text{ kde } t = 1, \dots, T.$$

Tento rozklad zachovává jednotky, takže po provedení rozkladu jsou všechny jeho složky vyjádřeny ve stejných jednotkách jako časová řada, ze které složky vycházejí.

- Multiplikativní rozklad znamená, že pro zpětné získání původní časové řady, je nutné mezi sebou všechny jednotlivé složky vynásobit:

$$y_t = T_t \cdot C_t \cdot S_t \cdot a_t, \text{ kde } t = 1, \dots, T.$$

Tato forma zachovává jednotky jen u trendové složky, ostatní jsou po dekompozici vyjádřeny relativně.²⁷

3.1 Modelování trendu

Tato kapitola se bude věnovat možnostem určení trendové složky. Nejprve je třeba nalézt funkci, která by odpovídala průběhu časové řady (viz kapitola 3.2). Metoda popisující trendovou složku pomocí funkce umožňuje snadnější předpovídání budoucích hodnot, protože parametry funkcí se nemění. Je však třeba poznamenat, že to ale pro delší časové období není reálné, proto se využívají další metody. K předpovídání slouží např. klouzavé průměry nebo exponenciální vyrovnávání.

Metodu modelování pomocí trendové funkce lze použít, když řada vyhovuje několika předpokladům. Při použití této metody jsou hodnoty časové řady prokládány jedinou funkcí. To může být u velkého množství řad velmi obtížné. Prvním z předpokladů je, že intervaly mezi jednotlivými měřeními musí být konstantní. Druhý říká, že časová řada má být vyjádřena jako $y_t = Y_t + a_t$, kde $t = 1, 2, \dots, T$, a Y_t zastupuje systematickou složku. Pokud je modelován trend, je systematická složka rovna trendové $Y_t = T_t$, a model časové řady má tvar $y_t = T_t + a_t$. Nesystematická složka a_t má vlastnosti bílého šumu. Jestli je zvolená funkce skutečně vhodná, ověřují různá kritéria (viz kapitola 3.3).²⁸

²⁷ Srov. HINDLS, Richard. *Statistika pro ekonomy*, s. 254-255.

²⁸ Srov. ARLT, Josef, Markéta ARLTOVÁ a Eva RUBLÍKOVÁ. *Analýza ekonomických časových řad s příklady*, s. 21.

3.2 Trendové funkce a jejich výběr

Mezi nejčastěji používané trendové funkce patří - funkce konstantní, lineární, kvadratická, exponenciální, S-křivka, modifikovaný exponenciální trend, logistický a Gompertzův trend. Modely v softwaru Mathematica hledá např. příkaz *Fit* jako lineární kombinaci zvolených funkcí. Ten využívá metody nejmenších čtverců (viz kapitola 3.2.1). Metodu nejmenších čtverců lze pro hledání parametrů využít u většiny z uvedených trendových funkcí. U posledních tří je nutné použít jiný způsob. Další možností je použít příkaz *FindFit*, který hledá nejvhodnější parametry pro zvolený model. Pro lineární modely je vhodný také příkaz *LinearModelFit* a pro nelineární *NonlinearModelFit*. Díky posledním dvěma příkazům lze jednoduše nalézt také rezidua (viz kapitola 3.3).²⁹

- Konstantní trend – hodnota ukazatele zůstává v čase stejná, tzn. $T_t = \beta_0$. Parametr β_0 se odhaduje pomocí metody nejmenších čtverců (viz kapitola 3.2.1). Platí $\hat{\beta}_0 = \frac{1}{T} \sum_{t=1}^T y_t$. Trend se tedy odhaduje jako aritmetický průměr hodnot časové řady.
- Lineární trend – vývoj hodnot kopíruje tvar přímky $T_t = \beta_0 + \beta_1 t$.
- Kvadratický trend – hodnoty se vyvíjí jako parabola $T_t = \beta_0 + \beta_1 t + \beta_2 t^2$.
- Exponenciální trend – exponenciální trendová funkce má tvar $T_t = \beta_0 \beta_1^t$. Metoda nejmenších čtverců pro odhady parametrů lze použít až po logaritmické transformaci ($\ln T_t = \ln \beta_0 + t \ln \beta_1$).
- S-křivka – po provedení transformace funkce logaritmováním se trend jeví jako hyperbola a má formu $T_t = e^{\beta_0 + \beta_1 \frac{1}{t}}$.
- Modifikovaný exponenciální trend – pro modifikovaný exponenciální trend platí $T_t = \gamma + \beta_0 \beta_1^t$, kde $\beta_0 \in (-\infty, 0)$, $\beta_1 \in (0, 1)$ a $\gamma \in (0, \infty)$. Jak je patrné, vychází z exponenciálního vývoje a je k němu přičtena konstanta γ . Tato konstanta představuje asymptotu, ke které funkce s rostoucím časem t konverguje. Takový trend nachází

²⁹ Srov. Wolfram Language & System Documentation Center. *Wolfram: Computation Meets Knowledge* [online]. Wolfram, 2020 [cit. 2020-03-14]. Dostupné z: <https://reference.wolfram.com/language/>.

velké využití v marketingu díky svému omezení. Protože funkci není možné zlinearizovat, pro zjištění odhadů parametrů se používají jiné metody.

- Logistický trend – popisuje jej Pearl-Reedova trendová funkce $T_t = \frac{1}{\gamma + \beta_0 \beta_1^t}$. Je tedy inverzní k předchozímu, modifikovanému exponenciálnímu, trendu.
- Gompertzův trend – popisuje jej funkce $T_t = \gamma \beta_0^{\beta_1^t}$ vychází stejně jako předchozí typ z modifikovaného exponenciálního trendu a jeho podobu získává po logaritmické transformaci. Odhady parametrů je možné zjistit stejně jako u modifikovaného exponenciálního trendu. Křivka se s rostoucím časem přibližuje zdola k asymptotě.³⁰

Prvotní výběr trendové funkce je možné provést na základě grafické analýzy. Funkce se volí přímo podle grafu časové řady nebo také pomocí grafické analýzy diferencí či koeficientů růstu (*Tabulka 1*).³¹

Tabulka 1 Volba trendu na základě hodnot 1. a 2. diferencí

Zdroj: vlastní zpracování

Trend	1. diference	2. diference
Konstantní	kolem 0	-
Lineární	kolem konstanty, která $\neq 0$	-
Kvadratický	mají lineární trend	kolem konstanty
Exponenciální	pokud řada koeficientů růstu nebo prvních diferencí $\ln y_t - \ln y_{t-1}$ kolísá kolem konstanty, která $\neq 0$	
S-křivka	pokud řada $\ln y_t$ kolísá kolem hyperboly	
Modifikovaný exponenciální	podíly sousedních diferencí se pohybují kolem konstanty, která $\neq 0$	
Gompertzova křivka	podíly diferencí tvaru $\frac{\ln y_t - \ln y_{t-1}}{\ln y_{t-1} - \ln y_{t-2}}$ se pohybují kolem konstanty, která $\neq 0$	

³⁰ HANČLOVÁ, Jana a Lubor TVRDÝ. *Úvod do analýzy časových řad*, s. 16-18.

³¹ Srov. ARLT, Josef, Markéta ARLTOVÁ a Eva RUBLÍKOVÁ. *Analýza ekonomických časových řad s příklady*, s. 26-27.

Při hledání parametrů modelu je nejčastěji využívána metoda nejmenších čtverců, která se zakládá na minimalizaci rozptylu hodnot. Její použití vyžaduje, aby parametry měly lineární formu. Některé funkce však na takový tvar převést nelze a je nutné použít pro odhad jinou metodu.

SM – ukázka syntaxe příkazů

Differences[list,n]

viz 2.3

Fit[list,{f₁,..., f_n},{x,y,...}]

hledá model $a_1 f_1 + \dots + a_n f_n$ pro seznam dat *list* pro funkce $f_1 \dots f_n$ o proměnných $\{x,y,\dots\}$

FindFit[list,expr,pars,vars]

hledá numerické hodnoty parametrů *pars* zvoleného modelu *expr* pro seznam dat *list* jako funkci proměnných *vars*

LinearModelFit[{y₁,y₂,...},{f₁,f₂,...},x]

sestavuje lineární model tvaru $\beta_0 + \beta_1 f_1 + \beta_2 f_2 + \dots$ pro hodnoty y_1, y_2, \dots s proměnnou x

NonlinearModelFit[{y₁,y₂,...},form,{β₁,...},x]

sestavuje nelineární model ve tvaru *form* pro hodnoty y_1, y_2, \dots s proměnnou x a parametry $\beta_1 \dots$

3.3 Interpolační kritéria

Ověřování toho, zda zvolený model dostatečně odpovídá analyzované časové řadě, lze realizovat buď prostřednictvím interpolačních nebo extrapolačních kritérií. Interpolační kritéria jsou založena na zkoumání chování odchylek odhadnutých hodnot od skutečných. Jejich rozdílem vznikají rezidua $\hat{a}_t = y_t - \hat{y}_t$, která vyjadřují odhady náhodné složky. V programu Mathematica lze rezidua jednoduše získat vyhodnocením vstupu *název*[“FitResiduals“]. Tomu ale musí předcházet tvorba modelu příkazem *LinearModelFit* nebo *NonlinearModelFit*.³² Chování reziduí je podkladem pro přijetí či odmítnutí modelu a zkoumá se pomocí kritérií, ke kterým patří:

³² Srov. Wolfram Language & System Documentation Center. *Wolfram: Computation Meets Knowledge* [online]. Wolfram, 2020 [cit. 2020-03-14]. Dostupné z: <https://reference.wolfram.com/language/>.

- Míry přesnosti vyrovnávání a průměrné charakteristiky reziduí:
 - Průměrná chyba se počítá jako:

$$ME = \frac{1}{T} \sum_{t=1}^T \hat{a}_t.$$

Jak říká název průměrná chyba se počítá jako aritmetický průměr všech reziduí, k tomu poslouží příkaz *Mean*, kde argumentem budou rezidua.

- Průměrná čtvercová chyba neboli rozptyl má tvar:

$$MSE = \frac{1}{T} \sum_{t=1}^T \hat{a}_t^2.$$

Pro zjištění této hodnoty software Mathematica nabízí příkaz *Variance*.

- Průměrná absolutní chyba se počítá jako:

$$MAE = \frac{1}{T} \sum_{t=1}^T |\hat{a}_t|.$$

Průměrná absolutní chyba se získává použitím příkazu *MeanDeviation* nebo složením příkazů *Mean* a *Abs*.

Následující dva druhy chyb už program jednoduchými příkazy získat neumí a je nutné formule složit z více příkazů. Lze použít *Mean*, *Abs* v kombinaci s matematickými operátory.

- Průměrná procentuální chyba je rovna

$$MPE = \frac{1}{T} \sum_{t=1}^T \frac{\hat{a}_t}{y_t} \cdot 100.$$

- Průměrná absolutní procentuální chyba se rovná:

$$MAPE = \frac{1}{T} \sum_{t=1}^T \frac{|\hat{a}_t|}{y_t} \cdot 100.^{33}$$

- Index determinace R^2 a modifikovaný index R_M^2 – vztah pro výpočet indexu determinace má tvar:

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

Jeho hodnota může nabývat intervalu $R^2 = \langle 0,1 \rangle$. Ověřovaný model je podle tohoto kritéria tím vhodnější, čím víc se hodnota indexu determinace blíží číslu 1. Ve vypočteném výsledku se však odráží počet parametrů zkoumané trendové funkce. Existuje proto také modifikovaný index determinace, který počet parametrů nezohledňuje. Je definován vztahem:

$$R_M^2 = R^2 - \frac{(1 - R^2)(k - 1)}{T - k}.$$

kde k je určeno počtem parametrů trendové funkce.³⁴ Oba typy indexů se vyhodnotí po použití nástrojů [“*RSquared*“] a [“*AdjustedRSquared*“] příkazu *LinearModelFit* nebo *NonlinearModelFit*.³⁵

- Autokorelační test náhodné složky softwaru Mathematica – testuje, jestli jsou zkoumané hodnoty autokorelovány. Autokorelace vyjadřuje závislost hodnot určité veličiny v čase na sobě samé. V případě ověřování vhodnosti zvolených modelů

³³ Srov. HINDLS, Richard. *Statistika pro ekonomy*, s. 288-290.

³⁴ Srov. ARLT, Josef, Markéta ARLTOVÁ a Eva RUBLÍKOVÁ. *Analýza ekonomických časových řad s příklady*, s. 29.

³⁵ Srov. Wolfram Language & System Documentation Center. *Wolfram: Computation Meets Knowledge* [online]. Wolfram, 2020 [cit. 2020-03-14]. Dostupné z: <https://reference.wolfram.com/language/>.

zkoumá zpožděné náhodné složky. V této práci bude popsán test autokorelace nabízený softwarem Mathematica, který je založen na testování hypotézy o náhodnosti dat. Nulová hypotéza H_0 říká, že autokorelace $\rho_1 = \rho_2 = \dots = \rho_k = 0$, a alternativní hypotéza H_a , že alespoň jedno $H_i \neq 0$. Zamítnutí nulové hypotézy připouští, že zkoumané hodnoty nejsou náhodné. Pro vyhodnocení software Mathematica používá 2 kritéria. Prvním z nich je Ljung-Boxovo:

$$n \cdot (n + 2) \sum_{i=1}^k \frac{\rho_i^2}{n - i}.$$

Druhé kritérium je BoxPiercovo:

$$n \sum_{i=1}^k \rho_i^2.$$

Zde n je délka souboru dat a k vyjadřuje velikost zpoždění. Software Mathematica po zadání příkazu *AutocorrelationTest* počítá p -hodnoty. Pokud je tato hodnota nízká, náhodnost dat je nepravděpodobná.³⁶

- Reziiduální autokorelační funkce (ACF) – zkoumá autokorelaci nesystematické složky. Nejprve je sestaven graf, kde vodorovná osa představuje časová zpoždění a na svislé ose jsou vyneseny koeficienty autokorelace ρ_k . Pokud se zde nevyskytuje takový koeficient, který by přesáhl meze 95 % spolehlivosti $(\frac{-2}{\sqrt{T}}, \frac{2}{\sqrt{T}})$, lze říci, že reziiduální složka autokorelaci neobsahuje. Koeficienty autokorelace se počítají pomocí vztahu:

$$\rho_k = \frac{\sum_{t=k+1}^T \hat{a}_t \cdot \hat{a}_{t-k}}{\sum_{t=1}^T \hat{a}_t^2} \in \langle -1, 1 \rangle. \text{ } ^{37}$$

³⁶ Srov. Wolfram Language & System Documentation Center. *Wolfram: Computation Meets Knowledge* [online]. Wolfram, 2020 [cit. 2020-03-14]. Dostupné z: <https://reference.wolfram.com/language/>.

³⁷ Srov. ARLT, Josef, Markéta ARLTOVÁ a Eva RUBLÍKOVÁ. *Analýza ekonomických časových řad s příklady*, s. 29.

K získání koeficientů autokorelační funkce slouží příkaz *AbsoluteCorrelationFunction* a vstupní buňka s obsahem *model["ParameterConfidenceIntervals"]* pro získání intervalu spolehlivosti.³⁸

SM – ukázka syntaxe příkazů

LinearModelFit[{ y_1, y_2, \dots }, { f_1, f_2, \dots }, x] nebo *NonlinearModelFit*[[{ y_1, y_2, \dots }, *form*, { β_1, \dots }, x]
názevmodelu["FitResiduals"]
vypíše rezidua modelu

Mean[*list*]

viz 2.2

Variance[*list*]

počítá rozptyl seznamu *list*

MeanDeviation[*list*]

počítá průměrnou absolutní odchylku hodnot seznamu *list*

Abs[x]

dává absolutní hodnotu čísla x

CorrelationFunction[*list*, n]

sestrojí korelační funkci pro seznam *list* a počet členů n

AutocorrelationTest[*list*, k , *options*]

testuje autokorelaci dat ze seznamu *list* se zpožděním k a možností dalších voleb *options*

Další možností, jak ověřit, zda zvolený model vyhovuje průběhu časové řady, jsou extrapoláční kritéria. Jejich princip spočívá v rozdělení řady na 2 části délky T_1 a T_2 , kde $T_1 + T_2 = T$. Na základě vývoje první části řady dochází k výběru modelu, odhadu parametrů a aplikaci interpolačních kritérií. Na druhé části řady se následně testuje vhodnost vybraného modelu pomocí tzv. předpovědi ex-post neboli předpovědi známé skutečnosti. Jedná se o zpětnou prognózu, kde se předpovídá zpětný vývoj pomocí výše zmíněných měř přesnosti vyrovnávání a průměrných charakteristik reziduí.

³⁸ Srov. Wolfram Language & System Documentation Center. *Wolfram: Computation Meets Knowledge* [online]. Wolfram, 2020 [cit. 2020-03-14]. Dostupné z: <https://reference.wolfram.com/language/>.

Usuzuje se, že pokud model dobře předpovídá minulý vývoj, je vhodný i pro předpovídání ex-ante neboli pro předpovědi neznámých hodnot.³⁹

3.4 Klouzavé průměry

Metoda klouzavých průměrů se od předchozího prokládání trendové funkce liší tím, že časová řada není vyrovnávána najednou. Řada je zde rozdělena na kratší úseky a nahrazena průměry vypočítaných z hodnot těchto úseků. Tento proces tvoří novou časovou řadu, která je vyrovnanější. Takovému průměrování se říká klouzavé, protože se úsek posunuje vždy o jednu hodnotu dále, tzn. nejstarší hodnota se odebere a přidá se jedno novější pozorování.

Pro začátek je důležité stanovení délky úseku neboli klouzavé části, které může významně ovlivnit další postup. Je vhodné ji zvolit jako liché číslo $2m + 1$, kde $m \in \mathbb{N}$. Důležitým parametrem pak je řád klouzavých průměrů. Jedná se o řád polynomu, který jednotlivé části vyrovnává. Také je třeba pro výpočet klouzavých průměrů zadat váhy hodnot. Váhy mohou u všech hodnot úseku zůstat stejné nebo se mohou postupně zmenšovat v závislosti na vzdálenosti od prostřední hodnoty.

Průměrování se liší podle toho, jak dlouhá období data představují. U sezonních dat s čtvrtletně zadanými hodnotami lze počítat centrované klouzavé průměry KP v čase t :

$$KP_t = \frac{x_t + x_{t+1} + x_{t-1} + \frac{1}{2}x_{t+2} + \frac{1}{2}x_{t-2}}{4}.$$

Obdobně pro sezonní řady s měsíčně zadanými hodnotami se počítají centrované klouzavé průměry podle vzorce:

$$KP_t = \frac{x_t + \sum_{i=1}^5 (x_{t+i} + x_{t-i}) + \frac{1}{2}x_{t+6} + \frac{1}{2}x_{t-6}}{12}.$$

³⁹ Srov. ARLT, Josef, Markéta ARLTOVÁ a Eva RUBLÍKOVÁ. *Analýza ekonomických časových řad s příklady*, s. 30.

Nevýhodou klouzavých průměrů je, že při výpočtu několik počátečních a koncových hodnot odpadne. Jedná se o první a poslední dvě u čtvrtletních časových řad a o prvních a posledních šest hodnot u měsíčních časových řad.⁴⁰

Klouzavé průměry se používají, když je třeba očistit časovou řadu od sezónních vlivů. Pomocí klouzavých průměrů lze totiž analyzovat chování sezónní složky. V prvním kroku je nutné vypočítat řadu klouzavých průměrů, který specifikuje trend a cyklickou složku. Ve druhém kroku dochází k eliminaci trendové a cyklické složky. To se provádí, pokud se pracuje s multiplikatívním modelem, dělením dat právě klouzavým průměrem a získává se osamostatněná sezónní a náhodná složka. Sezónní index (zkráceně SI) se počítá pro každé období v roce zvlášť (např. pro druhé čtvrtletí) a vyjadřuje porovnání tohoto období s typickým obdobím v roce. Získá se jako aritmetický průměr sezónních náhodných složek pro příslušné období.

Sezónní index také umožňuje z řady eliminovat sezónní složku, a to tak, že jednotlivé členy časové řady se vydělí příslušnými sezónními indexy.⁴¹

Pro příklady očišťování od sezónních vlivů je stěžejní příkaz *MovingAverage* a práce se seznamy. V práci se seznamem je potřebný např. příkaz *Partition*, nebo také *Part*.⁴²

SM – ukázka syntaxe příkazů

`MovingAverage[list,{w1,w2,...,wr}`

počítá klouzavý průměr seznamu dat *list* s vahami členů klouzavé části w_1, w_2, \dots, w_r

3.5 Předpovědi v časových řadách

Předpovědi jsou jednou z hlavních funkcí časových řad. Konstruuji se na základě subjektivních názorů odborníků (tzv. kvalitativní předpovědi) nebo matematicko-statistických modelů (tzv. kvantitativní předpovědi). Extrapolacemi se rozumí kvantitativní předpovědi s několika vlastnostmi. Vznikají tím, že se prodlouží vývoj hodnot časové řady do budoucnosti s horizontem h a za jinak stejných podmínek. Tyto extrapolace se dělí na bodové a intervalové

⁴⁰ HINDLS, Richard. *Statistika pro ekonomy*, s. 294.

⁴¹ Srov. SIEGEL, Andrew F. *Practical Business Statistics*, s. 438-441.

⁴² Srov. Wolfram Language & System Documentation Center. *Wolfram: Computation Meets Knowledge* [online]. Wolfram, 2020 [cit. 2020-03-14]. Dostupné z: <https://reference.wolfram.com/language/>.

Bodové předpovědi nebo také „extrapolace ex ante“ jsou předpovědi určené do okamžiku $t = T + h$, kde T je časový bod, ve kterém se začíná předpovídat. Odhad tedy představuje konkrétní číslo, je nutné ale brát v potaz, že při předpovědi dochází k chybě. Bodová předpověď bývá ještě doplněna o $(1 - \alpha).100\%$ interval spolehlivosti, tzn. pásmem, ve kterém se předpovězené hodnoty pohybují s pravděpodobností $(1 - \alpha).100\%$.

Protože předpovědi nebyvají stoprocentně přesné, vzniká odchylka neboli chyba předpovědi. Ta bývá způsobená výběrem modelu nebo nepřesným odhadem parametrů modelu. Definiuje se jako rozdíl mezi skutečnou a odhadnutou hodnotou:

$$\hat{a}_{T+h} = y_{T+h} - \hat{y}_T(h).^{43}$$

⁴³ Srov. ARLT, Josef, Markéta ARLTOVÁ a Eva RUBLÍKOVÁ. *Analýza ekonomických časových řad s příklady*, s. 24-25.

4 Analýza časových řad na základě Boxovy-Jenkinsovy metodologie

4.1 Základní principy

Tato kapitola bude věnována dalšímu způsobu vytváření lineárních modelů založeném na stochastické koncepci, kterou využívá Boxova-Jenkinsova metodologie. Takový přístup je znatelně flexibilnější a také schopný rychleji reagovat na výrazné změny v procesu než postupy popsané v předchozím textu. Popisuje i časové řady, které je obtížné analyzovat dříve uvedenými metodami. Na rozdíl od metody dekompozice se zaměřuje na práci se složkou náhodnou.

Použití Boxovy-Jenkinsovy metodologie vyžaduje, aby řada byla alespoň slabě stacionární. Slabá stacionarita se vyznačuje konstantní střední hodnotou a rozptylem a podmínkou, že závislost dvou pozorování byla dána jen jejich časovou vzdáleností, ne umístěním v řadě. Dále se bude uvažovat jen slabá stacionarita.

Dalším důležitým pojmem je autokorelační funkce (dále ACF). Ta vyjadřuje míru lineární závislosti mezi dvěma veličinami. Korelace ale může být také způsobena veličinami, které se nacházejí mezi nimi. Korelaci očištěnou od jejich vlivu lze zjistit pomocí parciální autokorelační funkce (dále PACF). Speciálním případem ACF je autokovarianční funkce, která vznikne tak, že se od ACF odečte střední hodnota.⁴⁴ Autokorelační funkci v softwaru Mathematica vykresluje příkaz *CorrelationFunction* a parciální korelační funkci *PartialCorrelationFunction*.⁴⁵

SM – ukázka syntaxe příkazů

`CorrelationFunction[list,hspec]`

odhadne korelační funkci pro seznam *list* se zpožděním *hspec*

`PartialCorrelationFunction[list,hspec]`

odhadne parciální korelační funkci pro seznam *list* se zpožděním *hspec*

`AutocorrelationTest[list,k,options]`

viz 3.3

⁴⁴ Srov. CIPRA, Tomáš. *Analýza časových řad s aplikacemi v ekonomii*, s. 31.

⁴⁵ Srov. Wolfram Language & System Documentation Center. *Wolfram: Computation Meets Knowledge* [online]. Wolfram, 2020 [cit. 2020-03-14]. Dostupné z: <https://reference.wolfram.com/language/>.

4.2 Modely a jejich identifikace

Ačkoli je modelování z části subjektivní, provádí se podle daného postupu ve 3 krocích, a to identifikací modelu, odhadu parametrů a ověřování. V první fázi hledání modelu a jeho řádu je vhodné nejprve znázornit řadu graficky pro ověření stacionarity. V případě, že toto kritérium nesplňuje, provádí se stacionarizace diferencováním nebo jinými metodami.⁴⁶ V programu Mathematica lze pro tuto úpravu použít příkaz *Differences*.⁴⁷ Stejně tak, když řada nemá nulovou střední hodnotu, je třeba ji centrovat. To se provádí odečtením aritmetického průměru od všech členů Y_t . Následně dochází k výběru modelu.

Základním stavebním prvkem Box-Jenkinsovy metodologie jsou lineární procesy. „Lineární proces představuje nekonečnou řadu ve tvaru:

$$y_t = \varepsilon_t + \psi_1 \varepsilon_{t-1} + \psi_2 \varepsilon_{t-2} + \dots,$$

kde ε_i reprezentují bílý šum a ψ_j jsou parametry.⁴⁸ Lineární proces neobsahuje předpověditelnou složku. Větší význam ale mají konkrétní typy těchto procesů, tj. proces klouzavých průměrů, autoregresní proces, integrovaný proces a smíšené procesy.

Proces klouzavých průměrů řádu q $MA(q)$ je stacionární a je popsán modelem:

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

kde ε_t jsou složky bílého šumu a θ_j jsou parametry. Pomocí operátoru posunutí B , pro který platí $B^j y_t = y_{t-j}$, lze tento proces popsat jako:

$$y_t = (1 - \theta_1 B - \dots - \theta_q B^q) \varepsilon_t.⁴⁹$$

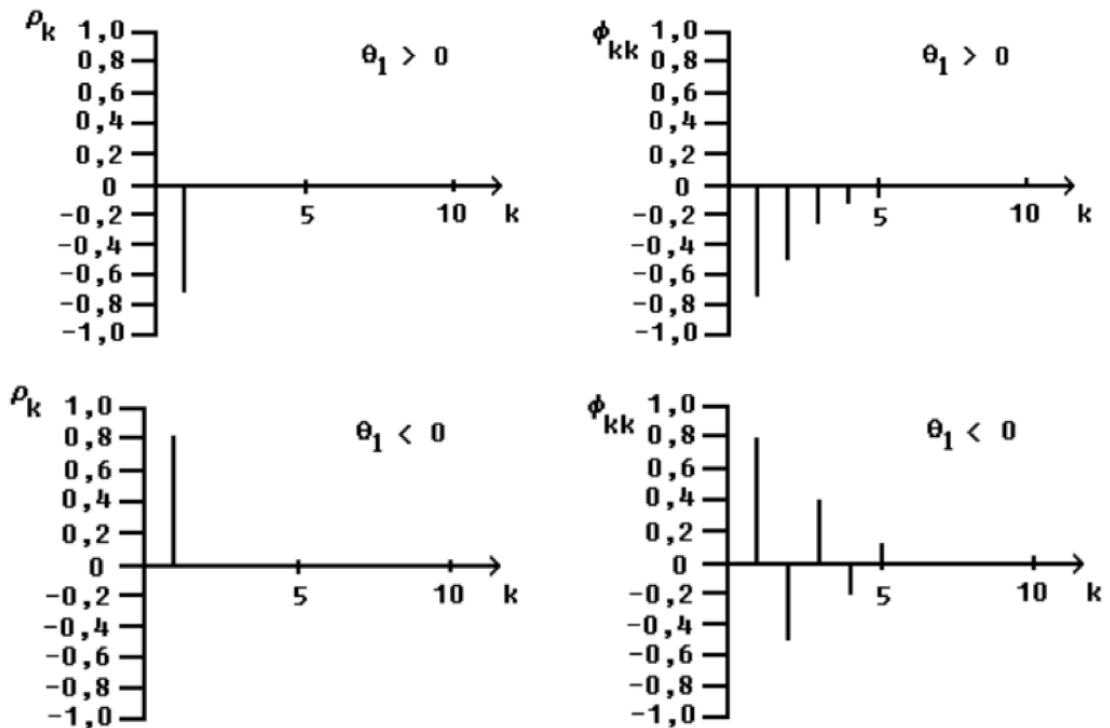
Následující obrázek zobrazuje charakteristické tvary pro autokorelační a parciální autokorelační funkci procesu klouzavých průměrů řádu 1 $MA(1)$ (Obrázek 6).

⁴⁶ Srov. CIPRA, Tomáš. *Analýza časových řad s aplikacemi v ekonomii*, s. 101.

⁴⁷ Srov. Wolfram Language & System Documentation Center. *Wolfram: Computation Meets Knowledge* [online]. Wolfram, 2020 [cit. 2020-03-14]. Dostupné z: <https://reference.wolfram.com/language/>.

⁴⁸ CIPRA, Tomáš. *Analýza časových řad s aplikacemi v ekonomii*, s. 107.

⁴⁹ Srov. ARLT, Josef, Markéta ARLTOVÁ a Eva RUBLÍKOVÁ. *Analýza ekonomických časových řad s příklady*, s. 92.



Obrázek 6 ACF a PACF modelu MA(1)

Zdroj: ARLT, Josef, Markéta ARLTOVÁ a Eva RUBLÍKOVÁ. *Analýza ekonomických časových řad s příklady*, s. 90.

Autoregresní proces řádu p $AR(p)$ lze vyjádřit jako:

$$y_t = \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + \varepsilon_t,$$

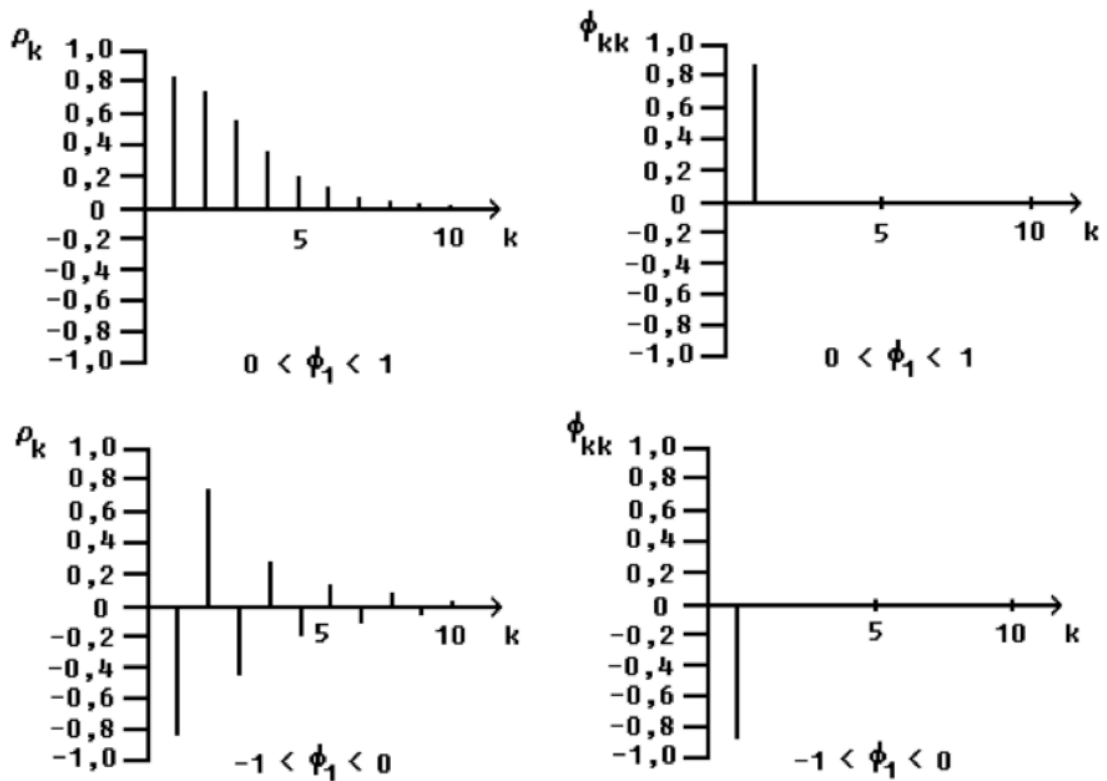
kde složky ϕ_p zastupují tzv. indikátory paměti. Čím se složky $|\phi_p|$ více blíží k jedné, tím je paměť delší. Pomocí operátoru zpětného posunutí lze proces vyjádřit jako:

$$(1 - \phi_1 B - \dots - \phi_p B^p) y_t = \varepsilon_t.$$

Tento proces je stacionární, jestliže kořeny rovnice $(1 - \phi_1 B - \dots - \phi_p B^p) = 0$ leží vně jednotkového kruhu.⁵⁰

⁵⁰ Srov. ARLT, Josef, Markéta ARLTOVÁ a Eva RUBLÍKOVÁ. *Analýza ekonomických časových řad s příklady*, s. 82.

Následující obrázek zobrazuje charakteristické tvary pro autokorelační a parciální autokorelační funkci autoregresního procesu řádu 1 $AR(1)$ (Obrázek 7).



Obrázek 7 ACF a PACF modelu $AR(1)$

Zdroj: ARLT, Josef, Markéta ARLTOVÁ a Eva RUBLÍKOVÁ. *Analýza ekonomických časových řad s příklady*, s. 87.

Smíšené autoregresní procesy klouzavých průměrů $ARMA(p, q)$ jsou kombinací předchozích dvou typů. Proces je vyjádřen ve tvaru:

$$y_t = \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q}.$$

Takový proces je stacionární, pokud je stacionární jeho autoregresní část. Stejně tak přebírá od autoregresního modelu tvar ACF.⁵¹ Identifikovat vhodný model lze na základě grafického vyjádření ACF a PACF (Obrázek 8).

⁵¹ Srov. CIPRA, Tomáš. *Analýza časových řad s aplikacemi v ekonomii*, s. 119.

Model	ACF	PACF
AR(p)	exponenciální a / nebo exponenciálně sinusoidní pokles	$\phi_{kk} = 0$ pro $k > p$
MA(q)	$\rho_k = 0$ pro $k > q$	Omezená exponenciálním a/nebo exponenciálně sinusoidním poklesem
ARMA(p, q)	Od zpoždění ($q - p$) pro $q > p$ exponenciální nebo exponenciálně sinusoidní pokles	Od zpoždění ($p - q$) pro $p > q$ omezená exponenciálním nebo exponenciálně sinusoidním poklesem

Obrázek 8 Tvary ACF a PACF pro modely AR, MA a ARMA

Zdroj: ARLT, Josef, Markéta ARLTOVÁ a Eva RUBLÍKOVÁ. *Analýza ekonomických časových řad s příklady*, s. 97.

Základem dalších nestacionárních procesů je proces náhodné procházky.

$$y_t = y_{t-1} + \varepsilon_t$$

Je vidět, že jeho první diference odpovídá procesu bílého šumu. Proto je nazýván integrovaným procesem 1. řádu a značí se $I(1)$. Pomocí operátoru zpoždění lze tento proces psát ve tvaru:

$$(1 - B)y_t = \varepsilon_t.$$

Integrovaný proces řádu d $I(d)$ je definován vztahem:

$$(1 - B)^d y_t = \varepsilon_t.^{52}$$

Autoregresní integrovaný proces klouzavých průměrů $ARIMA(p, d, q)$ je nestacionární proces tvaru:

$$(1 - \phi_1 B - \dots - \phi_p B^p)(1 - B)^d y_t = (1 - \theta_1 B - \dots - \theta_q B^q) \varepsilon_t$$

Pomocí transformace diferencování řádu d ho lze převést na stacionární invertibilní proces $ARMA(p, q)$.⁵³

⁵² Srov. ARLT, Josef a Markéta ARLTOVÁ. *Finanční časové řady*, s. 95.

⁵³ Srov. ARLT, Josef, Markéta ARLTOVÁ a Eva RUBLÍKOVÁ. *Analýza ekonomických časových řad s příklady*, s. 95.

Ověření vhodnosti modelu lze provádět stejně jako u metody dekompozice testem autokorelace.⁵⁴ Pro možnost jeho vyhodnocení je potřebné znát rezidua příslušného modelu. K jeho zjištění zde slouží tzv. Kalmanův filtr, který odhaduje trajektorii objektů na základě nepřesných pozorování. Pro jejich vyhodnocení v softwaru Mathematica slouží kombinace několika příkazů – *EstimatedProcess*, *Rest*, *KalmanFilter*, *Accumulate* a nástroj [“*PathStates*“] příkazu *TimeSeries*.⁵⁵

SM – ukázka syntaxe příkazů

Differences[list,n]

viz 2.3

TimeSeriesModelFit[list,mspec] + název[“*Process*“]

počítá koeficienty modelu zkonstruovaného příkazem *TimeSeriesModelFit*

EstimatedProcess[list,proc]

odhaduje pro hodnoty seznamu *list* koeficienty procesu *proc*

Rest[expr]

dává výraz *expr* bez prvního členu

TimeSeries[values,{times}] + název[“*PathStates*“]

dává trajektorii časové řady vytvořené příkazem *TimeSeries*

KalmanFilter[tproc,list]

zpřesňuje seznam *list* podle modelu časové řady *tproc*, kde prvním elementem výstupu je 0

Accumulate[list]

dává seznam postupně akumulovaných hodnot seznamu *list*

ARProcess[tproc,p]

sestrojí AR proces řádu *p* pro časovou řadu *tproc*

MAProcess[tproc,q]

sestrojí MA proces řádu *q* pro časovou řadu *tproc*

⁵⁴ Srov. KŘIVÝ, Ivan. *Analýza časových řad*, s. 62-63.

⁵⁵ Srov. Wolfram Language & System Documentation Center. *Wolfram: Computation Meets Knowledge* [online]. Wolfram, 2020 [cit. 2020-03-14]. Dostupné z: <https://reference.wolfram.com/language/>.

ARMAProcess[tproc,{p,q}]

sestrojí ARMA proces řádů p a q pro časovou řadu $tproc$

ARIMAProcess[tproc,{p,d,q}]

sestrojí ARIMA proces řádů p , d a q pro časovou řadu $tproc$

4.3 Prognózování pomocí modelu ARIMA

Odhadování budoucích hodnot pomocí Box-Jenkinsovy metodologie probíhá obdobně jako při predikování pomocí metody dekompozice. Nejprve je vybrán model ARIMA, který je následně ověřován použitím interpolačních kritérií. Pokud model testům vyhovuje, je vhodný i pro predikce.⁵⁶ Predikované hodnoty jsou lineární kombinací předchozích známých hodnot a je požadováno, aby předpověď měla minimální střední čtvercovou chybu.

Pro předpovědi neznámých hodnot časových řad se v programu Mathematica používá příkaz *TimeSeriesForecast*. Ten lze použít jen v případě, že je model konstruován pomocí příkazu *TimeSeriesModelFit*.⁵⁷

SM – ukázka syntaxe příkazů

TimeSeriesModelFit[list,mspec]

konstruuje model pro seznam dat *list* specifikovaný *mspec*

TimeSeriesForecast[tproc,list,k]

přiřadí seznamu hodnot *list* dle modelu *tproc* k -tou hodnotu po poslední známé hodnotě

⁵⁶ Srov. HINDLS, Richard. *Statistika pro ekonomy*, s. 330.

⁵⁷ Srov. Wolfram Language & System Documentation Center. *Wolfram: Computation Meets Knowledge* [online]. Wolfram, 2020 [cit. 2020-03-14]. Dostupné z: <https://reference.wolfram.com/language/>.

II. Praktická část

V této části práce bude postupně na konkrétních příkladech ukázáno použití softwaru Mathematica pomocí příkazů vysvětlených v teoretické části. Pro demonstraci budou analyzovány různé ekonomické řady z odvětví vědy a výzkumu. Většina zkoumání však bude zaměřena na jednu řadu, a to výdajů na výzkum a vývoj v České republice v letech 1989-2018. Bude pro ni počítán aritmetický průměr, míry dynamiky a také budou tvořeny modely jak metodou dekompozice tak Box-Jenkinsovou metodologií. Na závěr budou pro tuto řadu odhadovány budoucí hodnoty.

Pro některé ukázky budou však použity i jiné řady, všechna tato data, pokud se nejedná o časové řady generované programem, pocházejí ze záznamů Českého statistického úřadu.⁵⁸

Každý příklad se bude skládat ze zadání, popisu postupu a závěru. Budou rozebírány jak informace zjištěné o časových řadách tak i příkazy a formule sloužící pro vyhodnocování. Popsány budou vždy ty příkazy, které se ještě v předchozích příkladech neobjevily nebo jejich použití je odlišné.

⁵⁸ Český statistický úřad [online]. Praha, 2020 [cit. 2020-03-14]. Dostupné z: <https://www.czso.cz/csu/czso/domov>.

5 Získávání dat pro práci v programu Mathematica

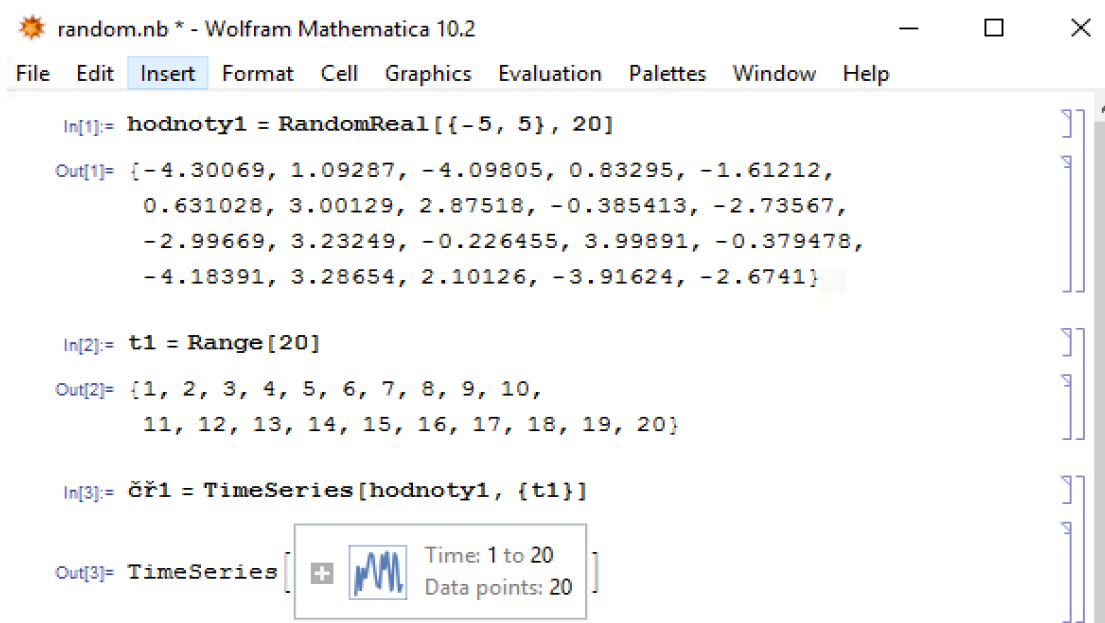
Příklady této části práce budou zaměřeny na generování, importování, a exportování dat. První z nich bude demonstrovat použití příkazů *RandomReal* a *RandomInteger* pro generování dat k tvorbě časové řady. Nakonec bude druhá z těchto časových řad exportována do jiného souboru.

Data je nutné často vkládat ze souboru, který má odlišný formát, ale není žádoucí je kopírovat či přepisovat. Bude proto uveden také příklad, který ukáže použití příkazu *Import*. Ten zajistí převod dat z jednoho formátu do druhého.

Příklad 5.1 (Generování vlastních dat)

Úkolem bude sestavit 2 časové řady, které budou mít 20 hodnot v intervalu $\langle -5; 5 \rangle$. První řada bude obsahovat hodnoty z množiny reálných čísel a časovou složku v podobě pořadí jednotlivých čísel. Druhá řada bude obsahovat hodnoty z množiny celých čísel a časovou složku ve formě let 2000-2019. Hodnoty druhé řady budou nakonec exportovány do souboru formátu *xls*.

Řešení: Nejprve budou vygenerovány náhodné hodnoty a časová složka, které pak budou propojeny a vznikne časová řada (viz kapitola 2.1). Následující obrázek ukazuje vytvoření první časové řady (Obrázek 9).



```
random.nb * - Wolfram Mathematica 10.2
File Edit Insert Format Cell Graphics Evaluation Palettes Window Help

In[1]:= hodnoty1 = RandomReal[{-5, 5}, 20]
Out[1]:= {-4.30069, 1.09287, -4.09805, 0.83295, -1.61212,
          0.631028, 3.00129, 2.87518, -0.385413, -2.73567,
          -2.99669, 3.23249, -0.226455, 3.99891, -0.379478,
          -4.18391, 3.28654, 2.10126, -3.91624, -2.6741}

In[2]:= t1 = Range[20]
Out[2]:= {1, 2, 3, 4, 5, 6, 7, 8, 9, 10,
          11, 12, 13, 14, 15, 16, 17, 18, 19, 20}

In[3]:= čř1 = TimeSeries[hodnoty1, {t1}]
Out[3]:= TimeSeries[Time: 1 to 20
                   Data points: 20]
```

Obrázek 9 Generování dat z množiny reálných čísel

Zdroj: vlastní zpracování

Ve vstupní buňce *In[1]* je příkazem *RandomReal* generován seznam 20 reálných hodnot z intervalu $(-5,5)$ a je pojmenován „*hodnoty1*“. Ve výstupní buňce *Out[1]* jsou vypsané hodnoty seznamu. Pro vytvoření časové řady je třeba doplnit hodnoty časové složky, k tomu bude použit příkaz *Range* ve druhé vstupní buňce. Tento vygeneruje posloupnost dvaceti celých čísel s diferencí 1. Postačí zadat horní mez, protože dolní je automaticky při jejím nevyplnění rovna 0. V dalším kroku bude vytvořena časová řada pomocí příkazu *TimeSeries*, jak je ukázáno ve třetí vstupní buňce.


Další obrázek obsahuje tvorbu druhé časové řady a její export (*Obrázek 10*). Postup v této ukázce je analogický jako při vytváření první časové řady. V první vstupní buňce dojde jen k jedné změně, a to ve tvaru příkazu, protože hodnoty mají pocházet z množiny celých čísel. Další změna nastane při sestavení časové složky, zde je nutné zadat koncovou, ale i počáteční hodnotu, přičemž diference zůstává stejná. Po té dojde k propojení seznamů příkazem *TimeSeries*, čímž vznikne časová řada.

```

random.nb * - Wolfram Mathematica 10.2
File Edit Insert Format Cell Graphics Evaluation Palettes Window Help

In[1]:= hodnoty2 = RandomInteger[{-5, 5}, 20]
t2 = Range[2000, 2019, 1];
čř2 = TimeSeries[hodnoty2, {t2}]

Out[1]= {2, 0, 3, -1, -5, -2, 2, -4, 1,
        -4, -5, -3, -1, 0, 1, -2, -5, -5, 1, 3}

Out[3]= TimeSeries [  Time: 2000 to 2019
                    Data points: 20 ]

In[4]:= seznamčř = {t2, hodnoty2};
Export["čř.xlsx", seznamčř, "Table"]

Out[5]= čř.xlsx

```

Obrázek 10 Generování dat z množiny celých čísel a jejich export

Zdroj: vlastní zpracování

V posledním kroku je řada exportována pomocí příkazu *Export*. Protože export dat se provádí ve formě seznamu, je nejprve spojena časová složka a hodnoty časové řady, to lze vidět v prvním řádku buňky *In[4]*. Koncový soubor obsahuje tabulku dat, je pojmenován zkratkou „*čř*“ a jeho formát je dán zkratkou *xls*.

Tyto ukázky demonstrovaly vytváření časových řad, rozdíly při používání příkazů *RandomReal* a *RandomInteger* a zadávání různých forem časových složek.

Příklad 5.2 (Import dat ze souboru)

Je dán soubor ve formátu *xls*, který obsahuje tabulku s daty o počtu registrovaných trestných činů v letech 2000 až 2018 v okrese Kladno (*Příloha 1*). Úkolem bude importovat tato data do notebooku softwaru Mathematica a pro větší přehlednost je uspořádat do tabulky.

Řešení: Importovaná data i výsledná tabulka jsou zobrazena v následujícím obrázku (*Obrázek 11*). V prvním řádku stojí příkaz *Import*, který přenese obsah souboru *ciny.xls* do notebooku softwaru Mathematica. Ve výstupní buňce je několikasložkový seznam, to je způsobeno argumentem *List* (viz kapitola 2.1) v předchozím příkazu *Import*.

```

In[1]:= Čř = Import["ciny.xls", "List"]
Out[2]= {{Období, 1, 2, 3, 4, 5, 6, 7},
  {Hodnota, 5541, 4923, 4916, 5034, 4859, 5024, 5537},
  {Období, 8, 9, 10, 11, 12, 13, 14},
  {Hodnota, 5624, 4610, 4237, 3878, 3754, 3677, 3934},
  {Období, 15, 16, 17, 18, 19}, {Hodnota, 3360, 3300, 3388, 3055, 2737}}

In[3]:= Čř1 = {"Období", 1, 2, 3, 4, 5, 6, 7},
  {"Hodnota", 5541, 4923, 4916, 5034, 4859, 5024, 5537}};
Čř2 = {"Období", 8, 9, 10, 11, 12, 13, 14},
  {"Hodnota", 5624, 4610, 4237, 3878, 3754, 3677, 3934}};
Čř3 = {"Období", 15, 16, 17, 18, 19},
  {"Hodnota", 3360, 3300, 3388, 3055, 2737}};

In[6]:= Grid[Čř1, Spacings -> {{2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2}, {0, .5}},
  Dividers -> {{{False, True, True, True, True, True, True, True}},
  {False, True, False}}]
Grid[Čř2, Spacings -> {{2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2}, {0, .5}},
  Dividers -> {{{False, True, True, True, True, True, True, True}},
  {False, True, False}}]
Grid[Čř3, Spacings -> {{2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2}, {0, .5}},
  Dividers -> {{{False, True, True, True, True, True}},
  {False, True, False}}]

Out[6]=


|         |      |      |      |      |      |      |      |
|---------|------|------|------|------|------|------|------|
| Období  | 1    | 2    | 3    | 4    | 5    | 6    | 7    |
| Hodnota | 5541 | 4923 | 4916 | 5034 | 4859 | 5024 | 5537 |



Out[7]=


|         |      |      |      |      |      |      |      |
|---------|------|------|------|------|------|------|------|
| Období  | 8    | 9    | 10   | 11   | 12   | 13   | 14   |
| Hodnota | 5624 | 4610 | 4237 | 3878 | 3754 | 3677 | 3934 |



Out[8]=


|         |      |      |      |      |      |
|---------|------|------|------|------|------|
| Období  | 15   | 16   | 17   | 18   | 19   |
| Hodnota | 3360 | 3300 | 3388 | 3055 | 2737 |


```

Obrázek 11 Import hodnot ze souboru

Zdroj: vlastní zpracování

Program umožňuje data vložit do tabulky přímo, a to vložením argumentu *Table* na místo argumentu *List*. Zde je tabulka vytvořena až po vložení dat, protože tento způsob zajišťuje větší adaptabilitu. Tabulky vzniklé pomocí příkazu *Grid* jsou stejně jako v původním souboru vytištěny do 3 částí, aby nevznikaly moc dlouhé řádky. Argumentem *Dividers* jsou navíc data oddělena čarami a argument *Spacing* umožňuje definování velikosti mezer mezi řádky a sloupci. Příklad ukázal, jakým způsobem je možné vložit data ze souboru formátu *xls* a jak z nich následně vytvořit přehlednou tabulku.

6 Základní charakteristiky časových řad

V této kapitole bude ukázáno, jakým způsobem lze vypočítat základní charakteristiky časové řady. Nejprve bude počítán průměr intervalové a okamžikové časové řady. U těchto řad se průměrná hodnota počítá různým způsobem. Výpočet prostého a chronologického průměru bude proto předveden na dvou různých řadách. Při výpočtu měr dynamiky však už není nutné rozlišovat mezi intervalovou a okamžikovou řadou, takže výpočet měr dynamiky lze demonstrovat už jen na jedné řadě.

Příklad 6.1 (Prostý aritmetický průměr intervalové časové řady)

Je dána časová řada výdajů na výzkum a vývoj v České republice v letech 1989-2018 (*Příloha 2*). Úkolem je zjistit, jaká je její průměrná hodnota.

Řešení: V tomto případě se jedná o intervalovou časovou řadu. Protože výdaje má smysl v průběhu intervalu sčítat, bude pro výpočet použit příkaz pro zjištění prostého aritmetického průměru (*viz kapitola 2.2*). Nejprve bude sestavena tabulka z hodnot časové řady.

První vstupní buňka v následujícím obrázku obsahuje pouze seznamy dat, proto není třeba vypisovat její výstupní buňku (*Obrázek 12*). Druhou vstupní buňkou *In[2]* je z vložených dat vytvořena tabulka se záhlavím příkazem *TableForm*. Dále je proveden výpočet prostého aritmetického průměru pomocí příkazu *Mean*. Jeho argument odkazuje na seznam hodnot, ze kterých se aritmetický průměr počítá. Výsledek se rovná zlomku $\frac{1277161}{30}$. Ten je dále numericky vyčíslen příkazem *N*. Průměrná hodnota této časové řady je rovna 42 472 mil. Kč.


```

průměr.nb * - Wolfram Mathematica 10.2
File Edit Insert Format Cell Graphics Evaluation Palettes Window Help

In[1]:= čř = {{1989, 1990, 1991, 1992, 1993, 1994, 1995, 1996},
             {16674, 12415, 15211, 14499, 12320, 12983, 13983, 16264}, {},
             {1997, 1998, 1999, 2000, 2001, 2002, 2003, 2004},
             {19477, 22865, 23647, 26487, 28337, 29552, 32247, 35083}, {},
             {2005, 2006, 2007, 2008, 2009, 2010, 2011, 2012},
             {38146, 43268, 50009, 49872, 50875, 52974, 62753, 72360}, {},
             {2013, 2014, 2015, 2016, 2017, 2018}, {77853, 85104, 88663, 80109, 90377, 102754}};

In[2]:= TableForm[čř,
                 TableHeadings ->
                 {"Rok", "Hodnota (mil. Kč)", "", "Rok", "Hodnota (mil. Kč)", "", "Rok",
                  "Hodnota (mil. Kč)", "", "Rok", "Hodnota (mil. Kč)", ""}]]

Out[2]/TableForm=
Rok      1989      1990      1991      1992      1993      1994      1995      1996
Hodnota (mil. Kč) 16674    12415    15211    14499    12320    12983    13983    16264

Rok      1997      1998      1999      2000      2001      2002      2003      2004
Hodnota (mil. Kč) 19477    22865    23647    26487    28337    29552    32247    35083

Rok      2005      2006      2007      2008      2009      2010      2011      2012
Hodnota (mil. Kč) 38146    43268    50009    49872    50875    52974    62753    72360

Rok      2013      2014      2015      2016      2017      2018
Hodnota (mil. Kč) 77853    85104    88663    80109    90377    102754

In[3]:= hodnoty = {16674, 12415, 15211, 14499, 12320, 12983, 13983, 16264, 19477, 22865,
                  23647, 26487, 28337, 29552, 32247, 35083, 38146, 43268, 50009, 49872, 50875,
                  52974, 62753, 72360, 77853, 85104, 88663, 80109, 90377, 102754};

In[4]:= Mean[hodnoty]

Out[4]=  $\frac{1277161}{30}$ 

In[5]:= N[%]

Out[5]= 42572.

```

Obrázek 12 Prostý aritmetický průměr intervalové časové řady

Zdroj: vlastní zpracování

Příklad 6.2 (Chronologický aritmetický průměr okamžikové časové řady)

Je dána řada počtu subjektů provádějících výzkum a vývoj v České republice v letech 1995-2018 (Příloha 3). I zde je úkolem spočítat její průměrnou hodnotu.

Řešení: Řada vyjadřuje počet subjektů zaznamenaný vždy v určitém okamžiku, proto se jedná o časovou řadu okamžikovou a průměrná hodnota se spočítá jako chronologický aritmetický průměr (viz kapitola 2.2). Výpočet je vidět na následujícím obrázku (Obrázek 13).

průměr2.nb * - Wolfram Mathematica 10.2

File Edit Insert Format Cell Graphics Evaluation Palettes Window Help

```

In[1]:= čř = {{1995, 1996, 1997, 1998, 1999, 2000, 2001, 2002},
             {570, 730, 807, 838, 1156, 1212, 1234, 1419}, {},
             {2003, 2004, 2005, 2006, 2007, 2008, 2009, 2010},
             {1692, 1795, 1855, 1966, 2021, 2047, 2155, 2392}, {},
             {2011, 2012, 2013, 2014, 2015, 2016, 2017, 2018},
             {2514, 2578, 2568, 2629, 2644, 2605, 2880, 2877}};

In[2]:= TableForm[čř,
                 TableHeadings ->
                 {"Rok", "Hodnota", "", "Rok", "Hodnota", "", "Rok", "Hodnota"}]

Out[2]/TableForm=


|         |      |      |      |      |      |      |      |      |
|---------|------|------|------|------|------|------|------|------|
| Rok     | 1995 | 1996 | 1997 | 1998 | 1999 | 2000 | 2001 | 2002 |
| Hodnota | 570  | 730  | 807  | 838  | 1156 | 1212 | 1234 | 1419 |
|         |      |      |      |      |      |      |      |      |
| Rok     | 2003 | 2004 | 2005 | 2006 | 2007 | 2008 | 2009 | 2010 |
| Hodnota | 1692 | 1795 | 1855 | 1966 | 2021 | 2047 | 2155 | 2392 |
|         |      |      |      |      |      |      |      |      |
| Rok     | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 |
| Hodnota | 2514 | 2578 | 2568 | 2629 | 2644 | 2605 | 2880 | 2877 |



In[3]:= hodnoty = {570, 730, 807, 838, 1156, 1212, 1234, 1419, 1692,
                  1795, 1855, 1966, 2021, 2047, 2155, 2392, 2514, 2578, 2568,
                  2629, 2644, 2605, 2880, 2877};

In[4]:= chronprům =
        Sum[(hodnoty[[i]] + hodnoty[[i + 1]]) / 2, {i, 1, Length[hodnoty] - 1}] /
        (Length[hodnoty] - 1)

Out[4]=  $\frac{86921}{46}$ 

In[5]:= N[%]

Out[5]= 1889.59

```

Obrázek 13 Aritmetický chronologický průměr okamžikové časové řady

Zdroj: vlastní zpracování

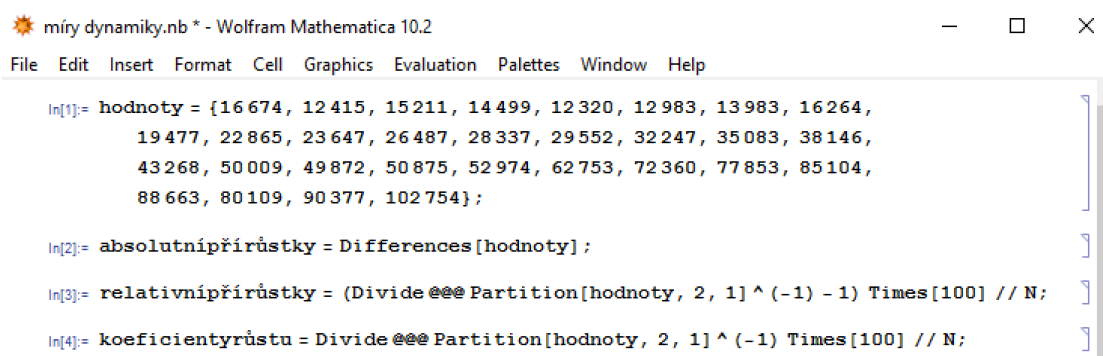
Tabulka zde byla tvořena analogicky jako v předchozím příkladu. Vlastní výpočet chronologického aritmetického průměru je zobrazen ve vstupní buňce *In[4]*. Sčítání ve výpočetní formuli je realizováno příkazem *Sum*. Poslední buňka *Out[5]* ukazuje, že vypočtený aritmetický chronologický průměr zadané okamžikové řady činí 1 889,59 subjektů.

Těmito dvěma příklady bylo ukázáno, jakým způsobem lze v softwaru Mathematica vypočítat průměr obou typů časových řad. Nyní lze přistoupit k výpočtu měr dynamiky.

Příklad 6.3 (Míry dynamiky)

Je dána řada výdajů na výzkum a vývoj v České republice v letech 1989-2018 (*Příloha 2*). Úkolem je spočítat míry dynamiky, tj. absolutní a relativní přírůstky, průměrný absolutní přírůstek, koeficient růstu a průměrný koeficient růstu.

Řešení: Následující tři obrázky ukazují, jak výpočet probíhá. První z nich bude obsahovat hodnoty a výpočetní formule pro absolutní a relativní přírůstky a koeficienty růstu. Druhý obrázek bude obsahovat tvorbu tabulky sestavenou z vypočtených hodnot a třetí ukáže výpočet průměrného absolutní přírůstku a průměrného koeficientu růstu.



```
míry dynamiky.nb * - Wolfram Mathematica 10.2
File Edit Insert Format Cell Graphics Evaluation Palettes Window Help

In[1]:= hodnoty = {16 674, 12 415, 15 211, 14 499, 12 320, 12 983, 13 983, 16 264,
19 477, 22 865, 23 647, 26 487, 28 337, 29 552, 32 247, 35 083, 38 146,
43 268, 50 009, 49 872, 50 875, 52 974, 62 753, 72 360, 77 853, 85 104,
88 663, 80 109, 90 377, 102 754};

In[2]:= absolutnípřírůstky = Differences[hodnoty];

In[3]:= relativnípřírůstky = (Divide@@@ Partition[hodnoty, 2, 1] ^ (-1) - 1) Times[100] // N;

In[4]:= koeficientyrůstu = Divide@@@ Partition[hodnoty, 2, 1] ^ (-1) Times[100] // N;
```

Obrázek 14 Formule pro absolutní a relativní přírůstky a koeficienty růstu

Zdroj: vlastní zpracování

Na obrázku (*Obrázek 14*) jsou postupně uvedeny 3 formule pro výpočet měr dynamiky (*viz kapitola 2.3*). Pro výpočet absolutních přírůstků slouží jednoduchý příkaz *Differences*. Výpočet relativních přírůstků je už ale složitější a skládá se z několika příkazů. *Partition* tvoří podseznamy vždy o délce 2, což zajistí druhý argument toho příkazu. Třetí argument, číslo 1, určuje získávání jednotlivých podmnožin posunem o jednu hodnotu dále. Dále je zapotřebí hodnoty uvnitř každé podmnožiny seznamu mezi sebou vydělit. Protože tato míra vyžaduje dělení každé hodnoty hodnotou předchozí, je nutné provést inverzi pořadí členů, a to umocněním na -1. Dále už jen stačí matematickými operátory nebo jednoduchými příkazy odečíst od všech výsledků číslo 1 a vynásobit číslem 100. Formule pro výpočet koeficientů růstu je uvedena v posledním řádku obrázku. Tvoří se analogicky jako pro výpočet předchozí míry dynamiky jen s tím rozdílem, že v závěru nedojde k odečtení čísla 1. Výsledky budou uvedeny v následujícím obrázku (*Obrázek 15*).

Tabulka je vytvořena ze všech vypočtených hodnot, pro její záhlaví bylo nutné navíc vytvořit časovou složku. Pomocí argumentu *TableAlignments* jsou všechna čísla zarovnaná

doprava a pomocí *TableDirection* nejsou seznamy hodnot vypsané v řádcích, ale ve sloupcích. V tabulce lze např. vidět, že v roce 1990 byly výdaje na výzkum a vývoj v České republice o 4 259 milionů Kč nižší než předchozí rok nebo že mezi lety 1989 a 1990 klesly výdaje o cca 25,5 %. Dalším údajem, který lze z tabulky vyčíst je, že výdaje klesly z roku 1989 na rok 1990 na 74,5 % a obdobně i u dalších let.

```

In[5]:= čas = {"1989-1990", "1990-1991", "1991-1992", "1992-1993", "1993-1994",
"1994-1995", "1995-1996", "1996-1997", "1997-1998", "1998-1999",
"1999-2000", "2000-2001", "2001-2002", "2002-2003", "2003-2004",
"2004-2005", "2005-2006", "2006-2007", "2007-2008", "2008-2009",
"2009-2010", "2010-2011", "2011-2012", "2012-2013", "2013-2014",
"2014-2015", "2015-2016", "2016-2017", "2017-2018"};
TableForm[{absolutnípřírůstky, relativnípřírůstky, koeficientyrůstu},
TableHeadings ->
{"absolutní přírůstky", "relativní přírůstky", "koeficienty růstu"}, čas},
TableAlignments -> {Right, Top}, TableDirections -> Row]

```

Out[8]/TableForm=

	absolutní přírůstky	relativní přírůstky	koeficienty růstu
1989-1990	-4259	-25.5428	74.4572
1990-1991	2796	22.5211	122.521
1991-1992	-712	-4.68082	95.3192
1992-1993	-2179	-15.0286	84.9714
1993-1994	663	5.38149	105.381
1994-1995	1000	7.70238	107.702
1995-1996	2281	16.3127	116.313
1996-1997	3213	19.7553	119.755
1997-1998	3388	17.3949	117.395
1998-1999	782	3.42007	103.42
1999-2000	2840	12.01	112.01
2000-2001	1850	6.98456	106.985
2001-2002	1215	4.28768	104.288
2002-2003	2695	9.11952	109.12
2003-2004	2836	8.79462	108.795
2004-2005	3063	8.73072	108.731
2005-2006	5122	13.4274	113.427
2006-2007	6741	15.5796	115.58
2007-2008	-137	-0.273951	99.726
2008-2009	1003	2.01115	102.011
2009-2010	2099	4.1258	104.126
2010-2011	9779	18.46	118.46
2011-2012	9607	15.3092	115.309
2012-2013	5493	7.59121	107.591
2013-2014	7251	9.31371	109.314
2014-2015	3559	4.18194	104.182
2015-2016	-8554	-9.64777	90.3522
2016-2017	10268	12.8175	112.818
2017-2018	12377	13.6949	113.695

Obrázek 15 Výsledky absolutních a relativních přírůstků a koeficientů růstu

Zdroj: vlastní zpracování

V dalším obrázku budou vypočteny zbývající dvě míry dynamiky (Obrázek 16). Průměrný absolutní přírůstek představuje aritmetický průměr z rozdílů vedlejší hodnot, proto postačí vložit příkaz *Differences* do příkazu *Mean*. Jeho výsledek je uveden ve výstupní buňce *Out[7]* a říká, že mezi lety 1989 a 2018 se v České republice výdaje na výzkum a vývoj ročně

zvyšovaly průměrně o 2 968,28 milionů Kč. Poslední mírou je průměrný koeficient růstu. Obrázek ukazuje jeho výpočet ve druhé vstupní buňce. Ve třetí vstupní buňce je od něj pro lepší interpretaci odečteno číslo 1 a následně je převeden na procentuální vyjádření. Jedním z důležitých příkazů, který počítá průměrný koeficient růstu, je příkaz *Product*. Ten násobí jednotlivé koeficienty růstu získané funkcí *hodnoty[[i+1]]/hodnoty[[i]]*, která každý člen vybraný ze seznamu hodnoty vydělí členem předchozím. Následně dojde k umocnění na počet členů snížený o 1. Z obrázku je patrné, že průměrný koeficient růstu je roven 1,064.

```

míry dynamiky.nb * - Wolfram Mathematica 10.2
File Edit Insert Format Cell Graphics Evaluation Palettes Window Help
In[7]:= průměrnýabsolutnípřírůstek = Mean[Differences[hodnoty]] // N
Out[7]= 2968.28

In[8]:= průměrnýkoeficientrůstu =
(Product[hodnoty[[i + 1]] / hodnoty[[i]], {i, 1, Length[hodnoty] - 1}) ^
(1 / (Length[hodnoty] - 1)) // N
Out[8]= {1.06471}

```

Obrázek 16 Průměrný absolutní přírůstek a průměrný koeficient růstu

Zdroj: vlastní zpracování

V příkladu bylo ukázáno, jak sestavit výpočetní formule pro zjištění hodnot měř dynamiky a jak z nich následně zkonstruovat přehlednou tabulku. Výsledné údaje byly nakonec interpretovány.

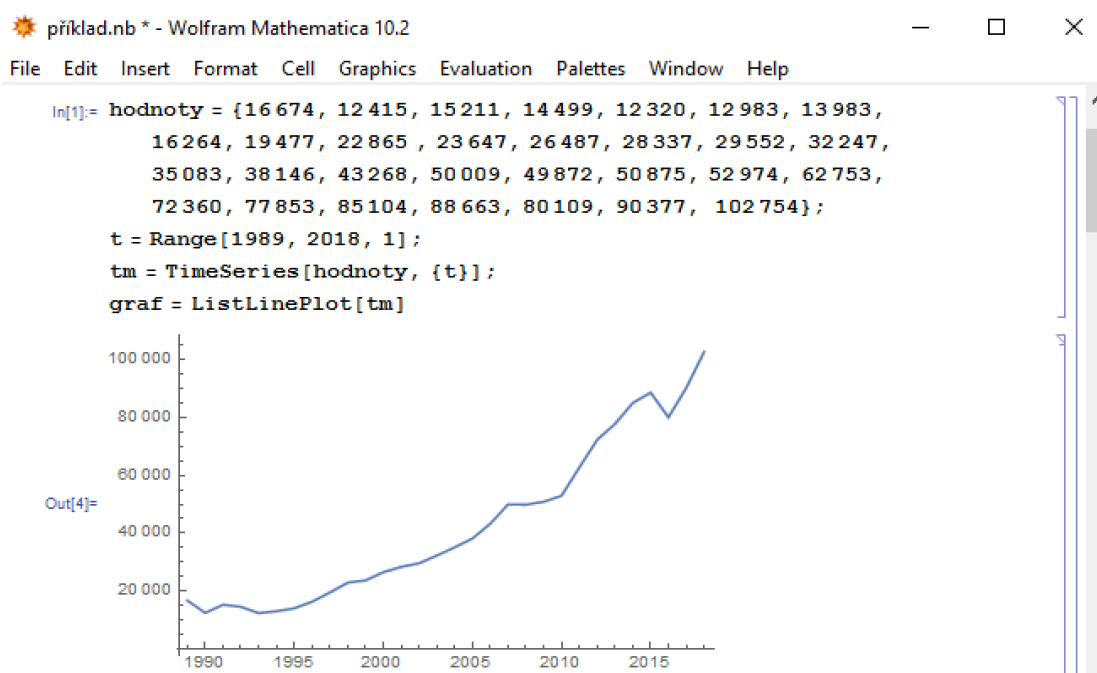
7 Metoda dekompozice

Modelování časové řady metodou dekompozice bude demonstrováno na jednom souhrnném příkladu. V něm bude na časové řadě předvedeno, jak a kterou trendovou funkci zvolit. Vybraná funkce pak bude porovnávána s další zvolenou trendovou funkcí.

Příklad 7.1 (Modelování trendů a jejich vzájemné porovnání)

Nechť je dána řada výdajů na výzkum a vývoj v České republice v letech 1989-2018 (Příloha 2). Úkolem bude na základě grafické analýzy vyhledat vhodnou trendovou funkci, pak zvolit jinou trendovou funkci a obě následně porovnat pomocí měr přesnosti, indexu determinace a modifikovaného indexu determinace.

Řešení: Nejprve bude provedena grafická analýza pro volbu vhodného trendu.

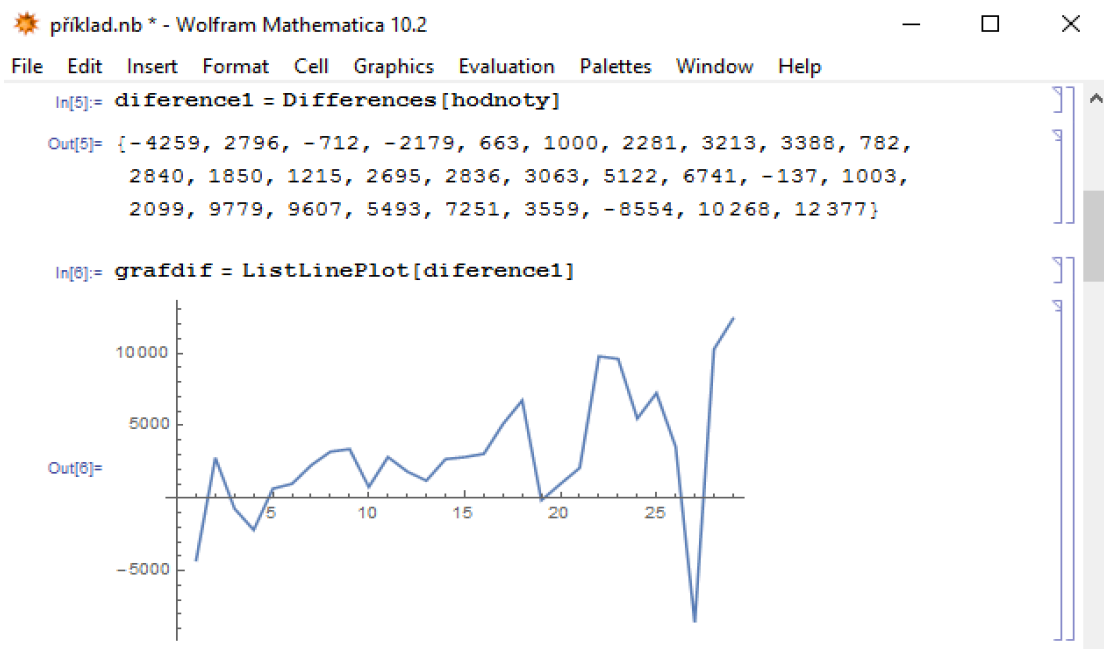


Obrázek 17 Vývoj hodnot časové řady

Zdroj: vlastní zpracování

V prvním kroku byla z dat vytvořena časová řada, která byla poté vykreslena do grafu pomocí *ListLinePlot* (Obrázek 17). Již podle grafu hodnot časové řady lze říci, že průběh se jeví jako kvadratický.

Tvrzení o kvadratickém modelu lze dle grafické analýzy ověřit prostřednictvím diferencí (viz kapitola 3.2). Řada prvních diferencí by se tedy měla pohybovat kolem lineárního trendu a druhé diference kolem konstantního trendu. V následujících obrázcích jsou zobrazeny jak ve formě seznamu, tak ve formě grafu (Obrázek 18-19). Časová složka bude dále u diferencí nahrazena pořadím hodnot.

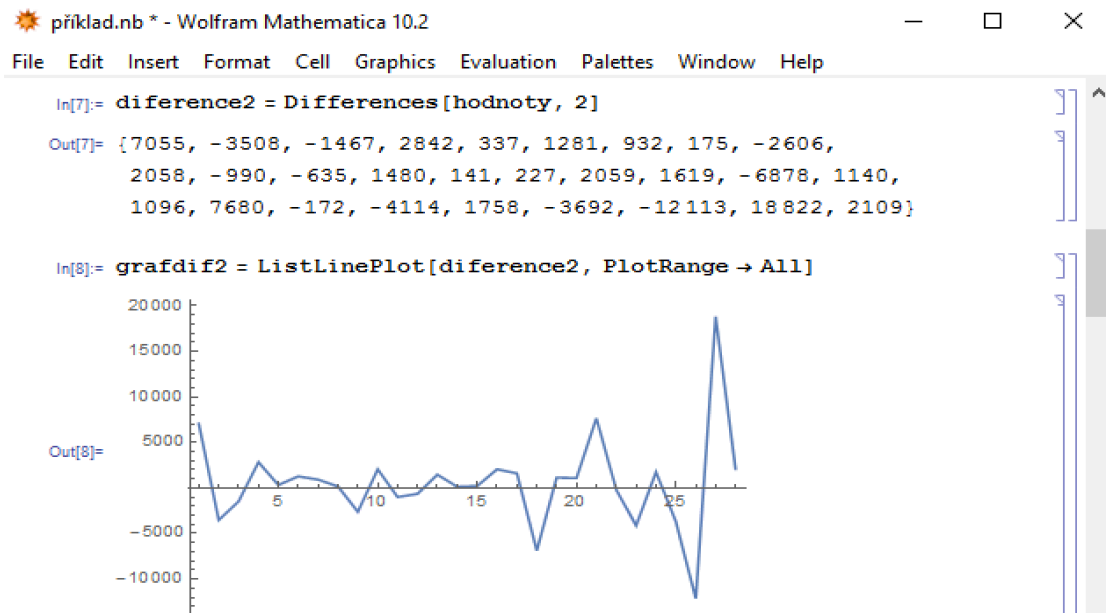


Obrázek 18 Průběh prvních diferencí časové řady

Zdroj: vlastní zpracování

Při vykreslení diferencí, které lze vidět v obrázku, lze soudit, že trend má přibližně lineární průběh (Obrázek 18). Je ale třeba vykreslit ještě graf druhých diferencí. Ten bude znázorněn v následujícím obrázku (Obrázek 19).

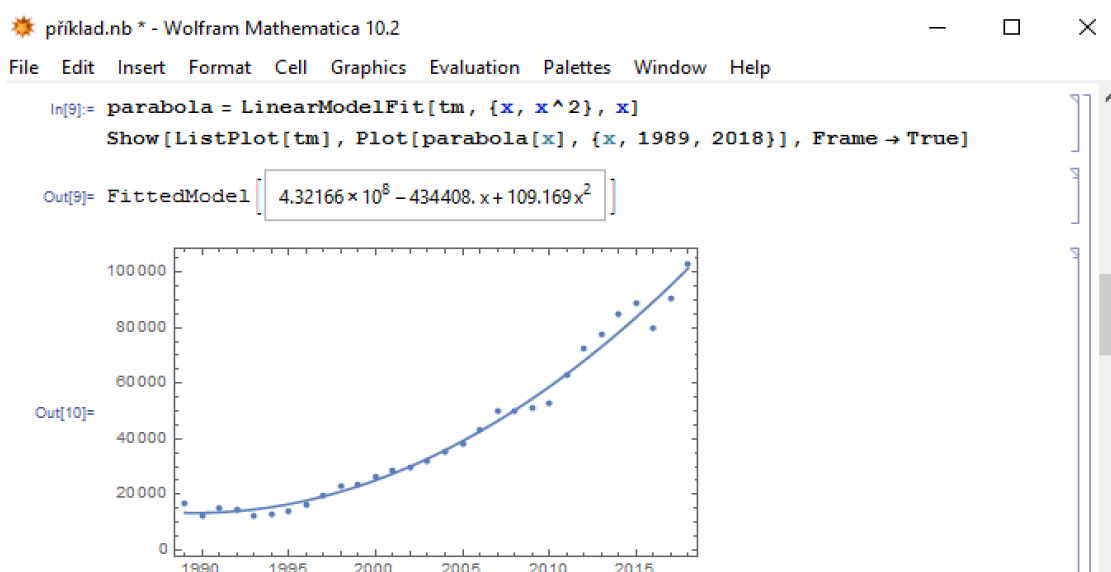
Obrázek ukazuje, že druhé diference se pohybují kolem konstantního trendu. Vývoj řady prvních a řady druhých diferencí naznačuje, že se jedná o kvadratický trend. Pro porovnávání je zvolena lineární trendová funkce.



Obrázek 19 Průběh druhých diferencí časových řad

Zdroj: vlastní zpracování

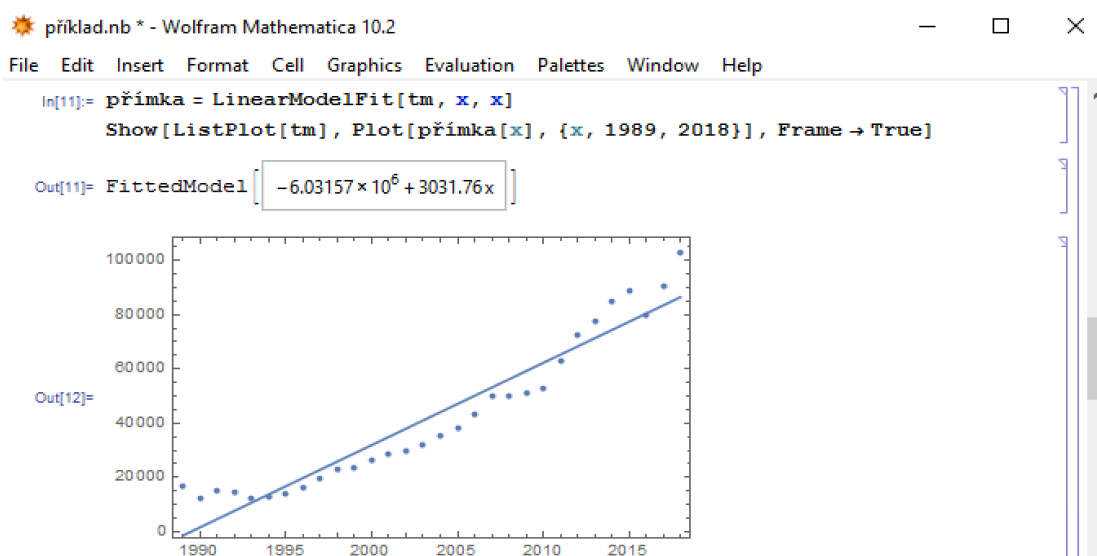
Dále budou modelovány určené trendy. Na následujícím obrázku je vidět kvadratický trend konstruovaný příkazem *LinearModelFit* (Obrázek 20). Tento příkaz bude vhodný pro následnou konstrukci reziduí. První argument příkazu udává časovou řadu, pro kterou je trend tvořen. Druhý říká, že ve funkci je jedna proměnná a trend bude kvadratický, a třetí udává proměnné funkce. Příkaz *Show* pak graf modelu i časové řady vykreslí a argument *Frame* vytvoří rámeček.



Obrázek 20 Kvadratický trend

Zdroj: vlastní zpracování

Analogicky je vytvořen i další obrázek se sestrojeným lineárním trendem (Obrázek 21). Modelování lineárního trendu se liší od kvadratického trendu odebráním kvadratického členu. Tato změna se projeví v rovnici i vykreslení trendové funkce, která tvoří parabolu. Jak bylo předpokládáno, rozdíly mezi hodnotami časové řady a trendové funkce jsou větší pro lineární model. Průběhu časové řady tedy lépe odpovídá kvadratická trendová funkce.

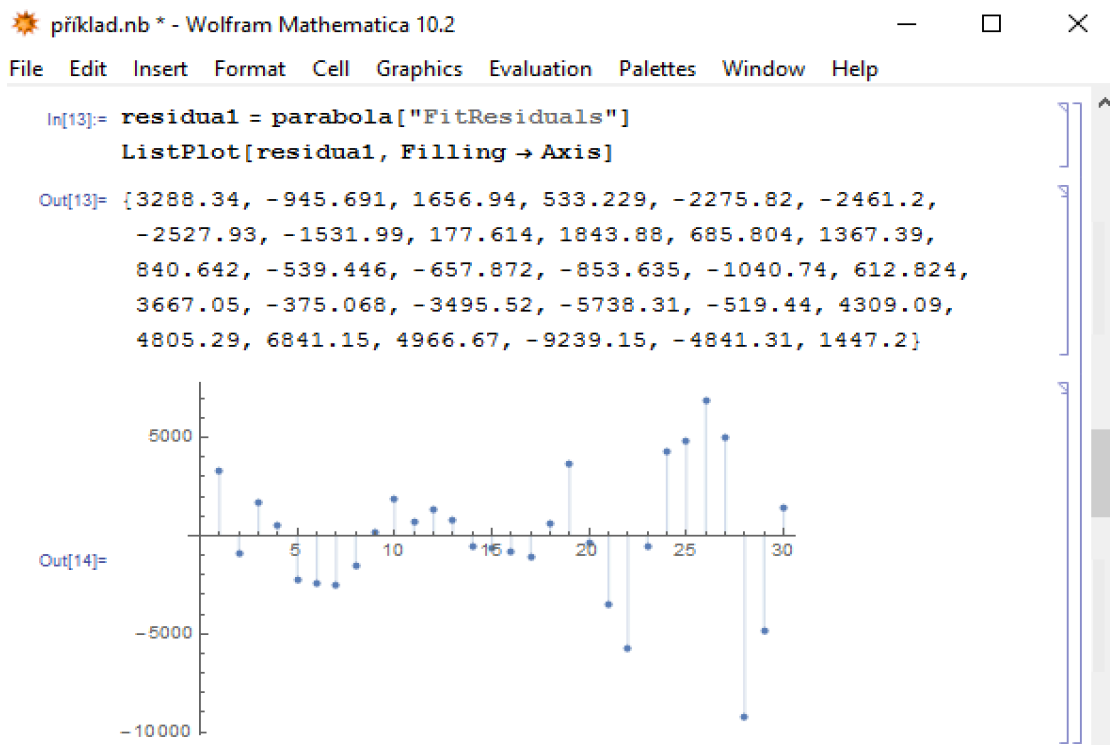


Obrázek 21 Lineární trend

Zdroj: vlastní zpracování

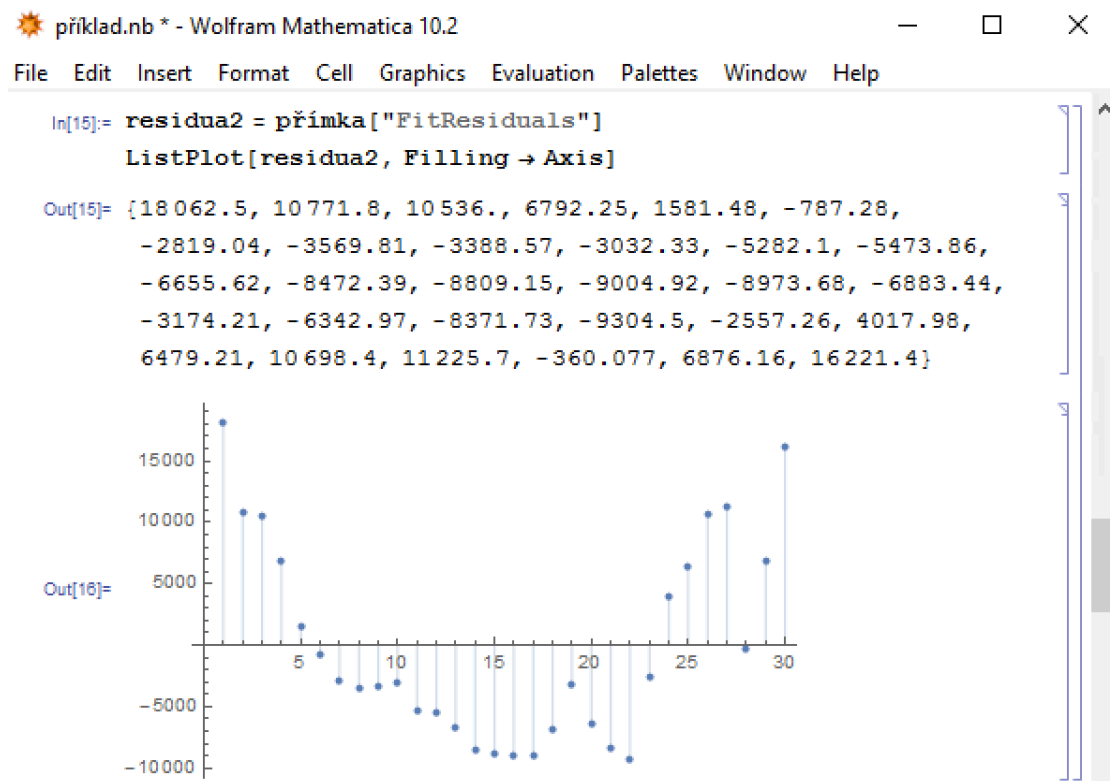
Dalším krokem analýzy řady bude ověřování vhodnosti trendů. Zkoumaným objektem budou tentokrát rezidua. Ta budou vyčíslena a vykreslena v následujících obrázcích (Obrázek 22-23). Na vodorovných osách grafu reziduí bude pro usnadnění manipulace zobrazeno jejich pořadí.

Protože byla trendová funkce konstruována příkazem *LinearModelFit*, lze rezidua přímo spočítat způsobem uvedeným ve vstupní buňce *In[13]*. Argument *FitResiduals* pak vyčíslí rezidua, která jsou následně příkazem *ListPlot* vykreslena do grafu. Pomocí argumentu *Filling* jsou vytvořeny spojnice rezidua s vodorovnou osou.



Obrázek 22 Rezidua kvadratického trendu

Zdroj: vlastní zpracování



Obrázek 23 Rezidua lineárního trendu

Zdroj: vlastní zpracování

Výše uvedená rezidua poslouží k porovnání měr přesnosti (viz kapitola 3.3) obou zvolených modelů. V následujícím obrázku je uveden výpočet těchto měr a jejich hodnoty uspořádané do tabulky (Obrázek 24). Pro výpočet prvních třech měr přesnosti postačují jednoduché příkazy *Mean*, *Variance* a *MeanDeviation*. Pro výpočet průměrné absolutní chyby lze použít složením příkazů *Abs* a *Mean*, kde příkaz *Abs* stanoví absolutní hodnotu ze všech reziduí a příkaz *Mean* následně vypočítá jejich aritmetický průměr. Tento postup byl použit v případě lineární trendové funkce.

```

príklad.nb * - Wolfram Mathematica 10.2
File Edit Insert Format Cell Graphics Evaluation Palettes Window Help
In[17]:= me1 = Mean[residua1];
mse1 = Variance[residua1];
mae1 = MeanDeviation[residua1];
mpe1 = Mean[residua1/hodnoty] * 100;
mape1 = Mean[Abs[residua1/hodnoty]] * 100;

In[22]:= me2 = Mean[residua2];
mse2 = Variance[residua2];
mae2 = Mean[Abs[residua2]];
mpe2 = Mean[residua2/hodnoty] * 100;
mape2 = Mean[Abs[residua2/hodnoty]] * 100;

In[27]:= TableForm[{{me1, mse1, mae1, mpe1, mape1}, {me2, mse2, mae2, mpe2, mape2}},
TableHeadings -> {"Kvadratický trend", "Lineární trend"},
{"ME", "MSE", "MAE", "MPE", "MAPE"}], TableAlignments -> {Right, Top}]
Out[27]/TableForm=

```

	ME	MSE	MAE	MPE	MAPE
Kvadratický trend	-1.83541×10^{-7}	1.12803×10^7	2469.54	-0.917412	6.93226
Lineární trend	1.3429×10^{-9}	6.6452×10^7	6884.2	2.08401	23.6855

Obrázek 24 Míry přesnosti kvadratického a lineárního trendu

Zdroj: vlastní zpracování

Průměrná procentuální chyba se počítá taktéž jednoduchým příkazem *Mean*, aplikovaného na stonásobek podílu hodnot reziduí a původních hodnot řady. Obdobně se počítá i průměrná absolutní procentuální chyba příkazem *Mean*, ten je však aplikován až po provedení absolutní hodnoty *Abs* na stejné podíly jako při výpočtu průměrné procentuální chyby.

Z tabulky na obrázku je vidět, že aritmetický průměr reziduí je v obou případech poměrně malý, průměrná čtvercová chyba je menší pro první trend, totéž platí pro průměrnou absolutní chybu, průměrnou procentuální chybu i pro absolutní procentuální chybu. Při porovnání velikost chyb pro první a druhý model vychází, že vhodnějším trendem je trend kvadratický.

Dalším zkoumaným jevem bude index determinace a modifikovaný index determinace (viz kapitola 3.3). Následující obrázek ukazuje, jakým způsobem lze tyto indexy vypočítat, a také jeho výsledky pro obě trendové funkce (Obrázek 25).

☀ příklad.nb * - Wolfram Mathematica 10.2

File Edit Insert Format Cell Graphics Evaluation Palettes Window Help

```

In[28]:= id1 = parabola["RSquared"];
mid1 = parabola["AdjustedRSquared"];
id2 = přímka["RSquared"];
mid2 = přímka["AdjustedRSquared"];

In[32]:= TableForm[{{id1, mid1}, {id2, mid2}},
  TableHeadings → {"Kvadratický trend", "Lineární trend"},
  {"Index determinace", "Modifikovaný index determinace"}],
  TableAlignments → {Right, Top}]

Out[32]/TableForm=

```

	Index determinace	Modifikovaný index determinace
Kvadratický trend	0.985516	0.984443
Lineární trend	0.914674	0.911627

Obrázek 25 Index a modifikovaný index determinace pro trendové funkce

Zdroj: vlastní zpracování

Index determinace a modifikovaný index determinace jsou nástroji příkazu *LinearModelFit*, proto jejich hodnoty lze v softwaru Mathematica získat obdobně jako hodnoty reziduí. Vhodnějším modelem je ten, který se více blíží číslu 1. Jak lze vidět na obrázku, tomuto kritériu vyhovuje lépe kvadratický trend (Obrázek 25).

Kvadratický trend se po celou dobu jevil jako vhodnější, to nakonec potvrdil i index determinace.

8 Klouzavé průměry

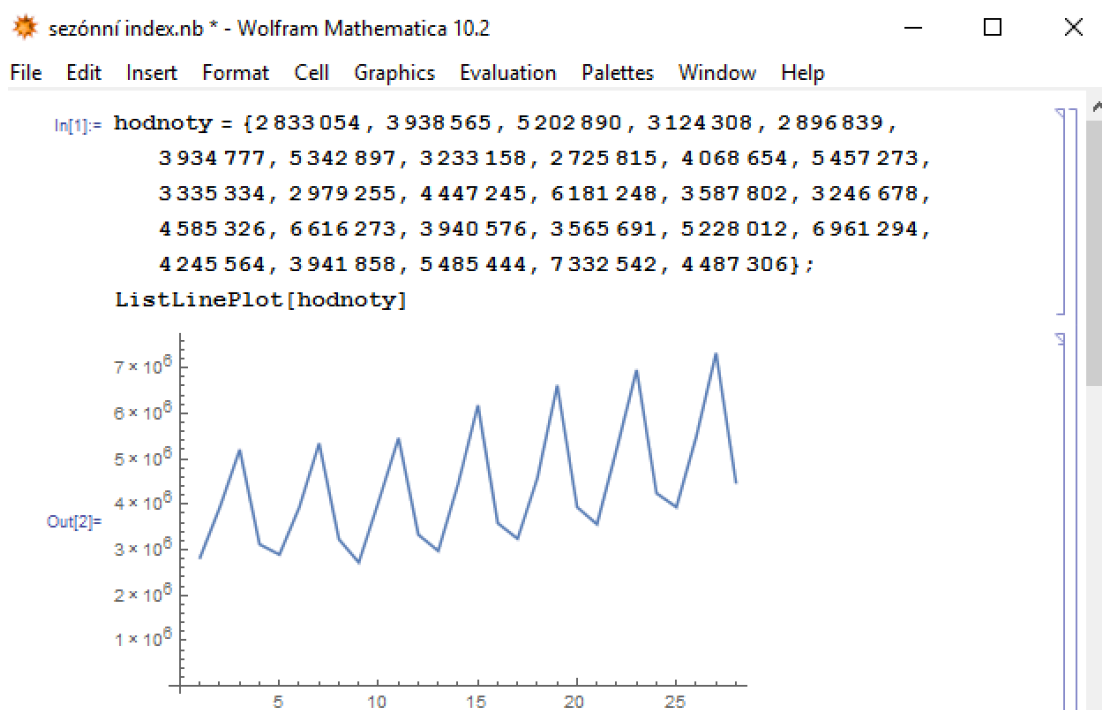
Následující příklad se zaměří na výpočet sezónního indexu, odstranění sezónní a náhodné složky a očištění řady od sezónních vlivů.

Příklad 8.1 (Očištění řady od sezónních vlivů)

Je dána čtvrtletní časová řada celkového počtu hostů ubytovacích zařízení České republiky v letech 2012-2018 (*Příloha 4*). Úkolem je očistit časovou řadu od sezónních vlivů.

Řešení: V příkladu bude nejprve klouzavými průměry eliminována trendová a cyklická složka časové řady. Poté bude osamostatněna sezónní a náhodná složka a pro každé období bude vypočten sezónní index. Nakonec bude časová řada očištěna od sezónní složky (*viz kapitola 3.4*).

Nejprve budou data vykreslena do grafu (*Obrázek 26*). Časová složka bude v tomto příkladu nahrazena pořadím hodnot, aby bylo možné provádět všechny operace.

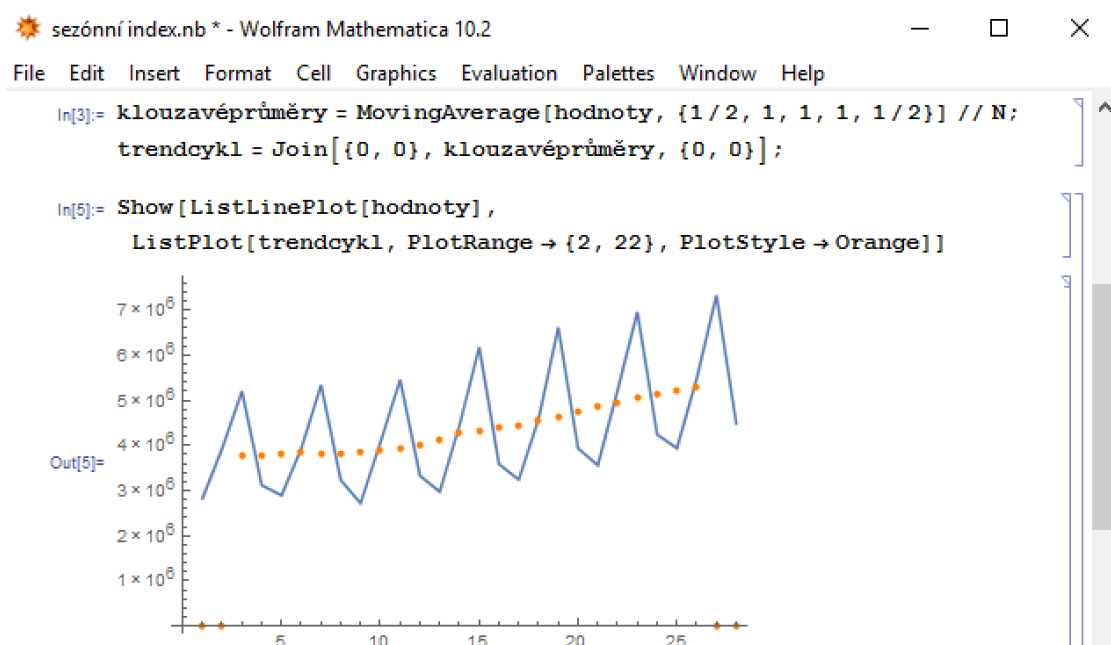


Obrázek 26 Vývoj hodnot časové řady

Zdroj: vlastní zpracování

V obrázku jsou nejprve vyobrazeny hodnoty čtvrtletní časové řady a následně je její vývoj znázorněn ve spojnicovém grafu. Graf poukazuje na výkyvy, ale vývoj má rostoucí trend.

V dalším kroku bude osamostatněna právě trendová složka spolu s cyklickou. Ve vstupní buňce *In[3]* obrázku lze vidět formuli pro zisk zmiňovaných dvou složek, dále je vytvořen graf pro porovnání původních hodnot časové řady. Výpočet je proveden průměrováním hodnot časové řady, což zajistí příkaz *MovingAverage*. Klouzavým částem je ale ještě také zapotřebí přiřadit váhy, a to tak, aby krajní hodnoty měly váhu poloviční. Délka klouzavé části je rovna 5, počet vah tedy musí být také roven tomuto číslu. Průměr se vztahuje vždy k prostřední hodnotě klouzavé části, proto je ve výsledném seznamu o 4 hodnoty méně. Výsledek tak chybí pro první a poslední dvě hodnoty. V dalším obrázku lze vidět, že graf je již mnohem vyhlazenější než před odstraněním sezónní a náhodné složky (*Obrázek 27*).



Obrázek 27 Porovnání hodnot časové řady a graf trendové a cyklické složky

Zdroj: vlastní zpracování

V dalším kroku budou zjišťovány koeficienty sezónní a náhodné složky, bude tak naopak eliminován trend a složka cyklická. Spočteny budou také sezónní indexy (*Obrázek 28*).

```

sezónní index.nb * - Wolfram Mathematica 10.2
File Edit Insert Format Cell Graphics Evaluation Palettes Window Help
In[6]:= hodnotysezónníindex = Take[hodnoty, {3, 26}];
In[7]:= sezónníanáhodnáložka = hodnotysezónníindex/klouzavéprůměry
Out[7]:= {1.37545, 0.824317, 0.760884, 1.02513, 1.39482, 0.845072,
0.706732, 1.04754, 1.38917, 0.832283, 0.718711, 1.04216,
1.42678, 0.818572, 0.72883, 1.00707, 1.42682, 0.828322,
0.73056, 1.05361, 1.37925, 0.828183, 0.757329, 1.0386}
In[8]:= rozdělenípodlelet = Partition[sezónníanáhodnáložka, 4];
si1 = Mean[rozdělenípodlelet[[All, 3]]];
si2 = Mean[rozdělenípodlelet[[All, 4]]];
si3 = Mean[rozdělenípodlelet[[All, 1]]];
si4 = Mean[rozdělenípodlelet[[All, 2]]];
In[13]:= TableForm[{{si1, si2, si3, si4}},
TableHeadings -> {"sezónní index", {"si1", "si2", "si3", "si4"}},
TableAlignments -> {Right}]
Out[13]/TableForm=

```

	si1	si2	si3	si4
sezónní index	0.733841	1.03569	1.39871	0.829458

Obrázek 28 Výpočet sezónní a náhodné složky a sezónního indexu

Zdroj: vlastní zpracování

Protože je počáteční seznam dělen získanými koeficienty, bylo nutné z něj nejprve odstranit první a poslední dvě hodnoty, aby si odpovídali příslušní dělenci a dělitelé. Odstranění bylo provedeno pomocí příkazu *Take*, který ze seznamu hodnot *hodnoty* vyjmul třetí až dvacátý šestý člen. Po výpočtu koeficientů lze přistoupit k výpočtu sezónních indexů *si1, ... si4* (viz kapitola 3.4), které poslouží k eliminaci náhodné složky. Jejich výpočet je zobrazen v buňce *In[8]*. Příkaz *Partition* po řadě rozdělí koeficienty do podseznamů, kde každý z nich bude mít 4 členy, tedy bude pokrývat vždy 4 čtvrtletí. Dvojitá závorka pak provede výběr vždy *n*-tého členu každého podseznamu. Např. u prvního sezónního indexu vybere každý třetí člen, protože výpočtem klouzavých průměrů byly první dvě hodnoty ze seznamu vyřazeny. Tyto vybrané členy jsou pak průměrovány příkazem *Mean*. Pro druhá, třetí o čtvrtá období jsou indexy tvořeny analogickým způsobem. Výsledky jsou následně zaznamenány do tabulky pomocí příkazu *TableForm*.

Když jsou vypočteny sezónní indexy, lze očistit časovou řadu od sezónních vlivů (Obrázek 29).

```

sezónní index.nb * - Wolfram Mathematica 10.2
File Edit Insert Format Cell Graphics Evaluation Palettes Window Help

In[14]:= čtvrtletníhodnoty = Partition[hodnoty, 4];
očištěnéprvníčtvrtletí = čtvrtletníhodnoty[[All, 1]] / si1;
očištěnédruhěčtvrtletí = čtvrtletníhodnoty[[All, 2]] / si2;
očištěnéřetíčtvrtletí = čtvrtletníhodnoty[[All, 3]] / si3;
očištěnéčtvrtéčtvrtletí = čtvrtletníhodnoty[[All, 4]] / si4;

In[19]:= TableForm[{očištěnéprvníčtvrtletí, očištěnédruhěčtvrtletí,
očištěnéřetíčtvrtletí, očištěnéčtvrtéčtvrtletí},
TableHeadings ->
{{"1.čtvrtletí", "2.čtvrtletí", "3.čtvrtletí", "4.čtvrtletí"},
{"2012", "2013", "2014", "2015", "2016", "2017", "2018"}},
TableDirections -> Column]

Out[19]/TableForm=

```

	2012	2013	2014	2015
1.čtvrtletí	3.86058×10^6	3.9475×10^6	3.71445×10^6	4.05981
2.čtvrtletí	3.80285×10^6	3.7992×10^6	3.92846×10^6	4.29401
3.čtvrtletí	3.71977×10^6	3.81987×10^6	3.90164×10^6	4.41924
4.čtvrtletí	3.76669×10^6	3.89792×10^6	4.0211×10^6	4.32548

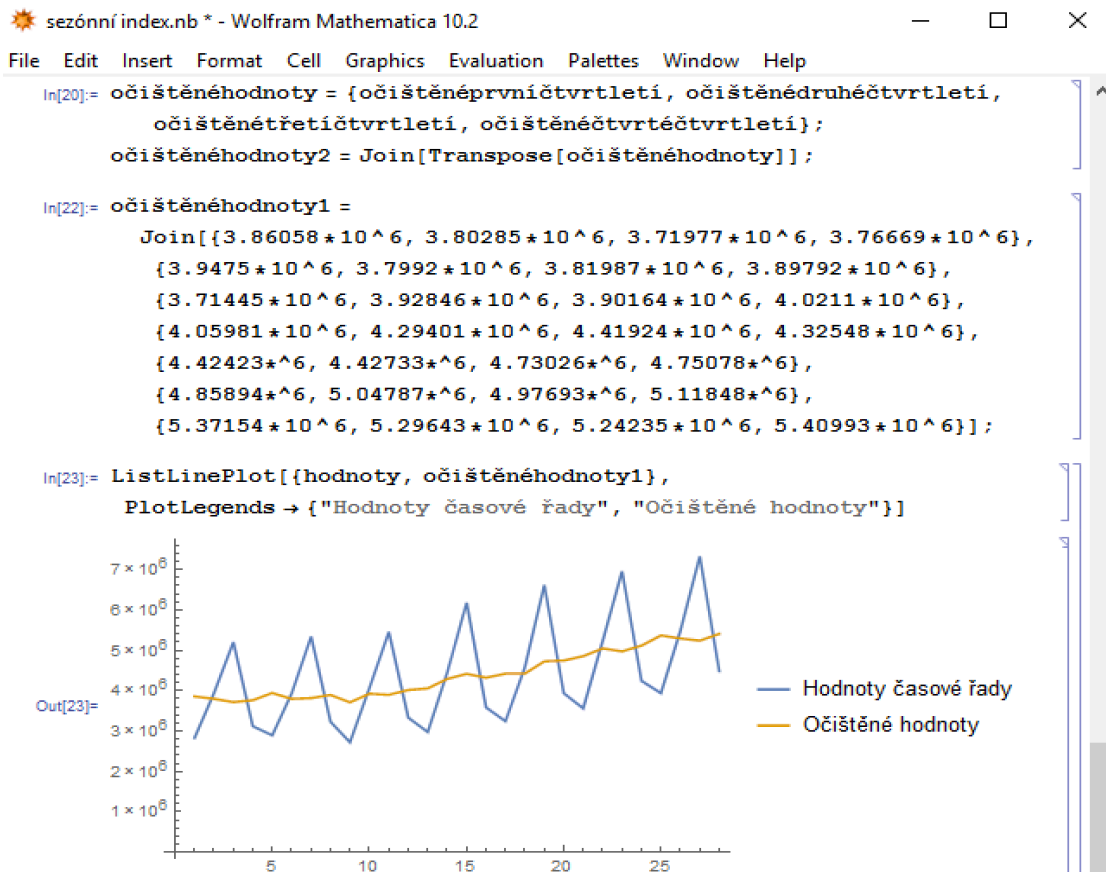
Obrázek 29 Hodnoty očištěné od sezónních vlivů

Zdroj: vlastní zpracování

Očištěné hodnoty lze získat pouhým vydělením hodnot příslušných čtvrtletí sezónním indexem. Předtím je ale nutné hodnoty časové řady po řadě rozdělit do čtveřic příkazem *Partition* a z těchto podseznamů pomocí dvojité hranaté závorky vybrat hodnoty určitých období. Očištěné hodnoty časové řady jsou uvedeny ve spodní části obrázku v tabulce.

V poslední fázi už bude jen vykreslen graf očištěných hodnot s původními hodnotami časové řady. První buňka obrázku ukazuje tvorbu jednoho velkého seznamu, který bude tvořen také seznamy, a to očištěných období. Hodnoty jsou prozatím seskupeny podle čtvrtletí, pro graf je ale potřebné seskupit je zpět podle času příkazem *Transpose*. Po použití příkazu *Join*, který je spojí v jeden seznam, jsou hodnoty připravené pro sestrojení grafu. Zobrazený graf znázorňuje vývoj hodnot časové řady spolu s očištěnými hodnotami (Obrázek 30).

Výstupem řešeného příkladu jsou hodnoty očištěné od sezónních vlivů, které jsou pro lepší názornost vykresleny spolu s původními hodnotami v grafu. Protože je graf vykreslen příkazem *ListLinePlot*, bylo možné k němu přidat legendu. Také bylo ukázáno, jak získat sezónní indexy, trendovou a cyklickou složku nebo také naopak sezónní a náhodnou složku.



Obrázek 30 Porovnání hodnot časové řady a hodnot očištěných od sezónních vlivů
 Zdroj: vlastní zpracování

9 Box-Jenkinsova metodologie

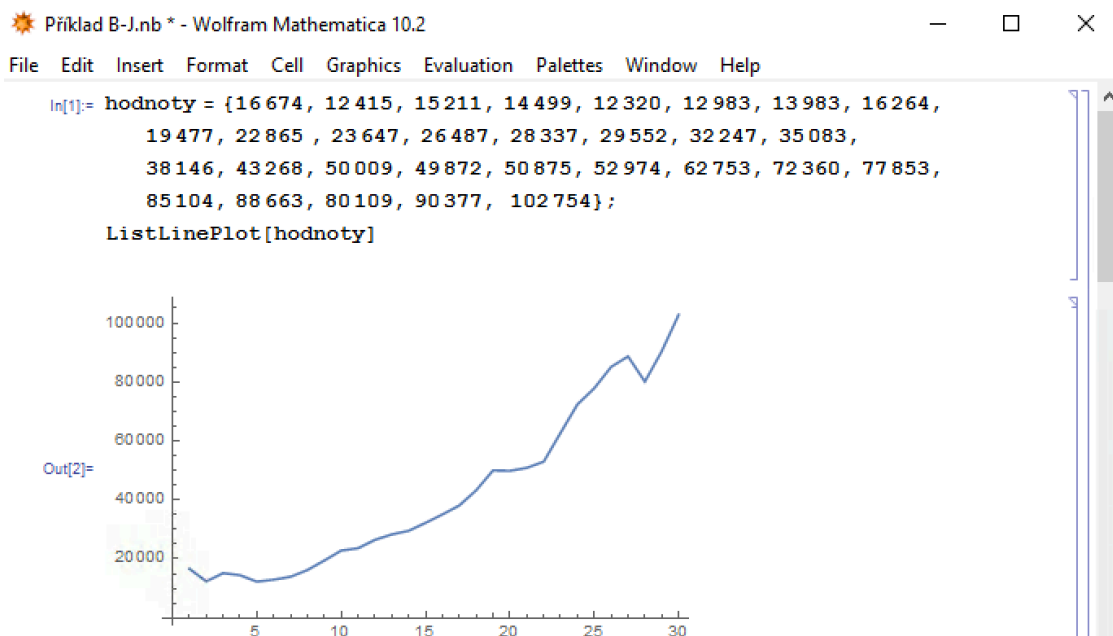
V této části bude uvedeny dva příklady. V prvním z nich bude modelována časová řada pomocí Box-Jenkinsovy metodologie. Bude ověřována stacionarita časové řady, pak software Mathematica vybere vhodný proces, který bude následně ověřován. V druhém příkladu budou pomocí vybraného modelu předpovídány budoucí hodnoty.

Příklad 9.1 (Konstrukce ARIMA modelu)

Je dána řada výdajů na výzkum a vývoj v České republice v letech 1989-2018. Úkolem je ověřit stacionaritu, vybrat ARIMA model a provést test autokorelace.

Řešení: Příklad bude řešen podle následujícího postupu. Nejprve bude pomocí softwaru Mathematica ověřováno, zda je řada stacionární, případně bude stacionarizována. Dále bude proveden výběr vhodného modelu a vyhodnocen test autokorelace (*viz kapitola 3.3*).

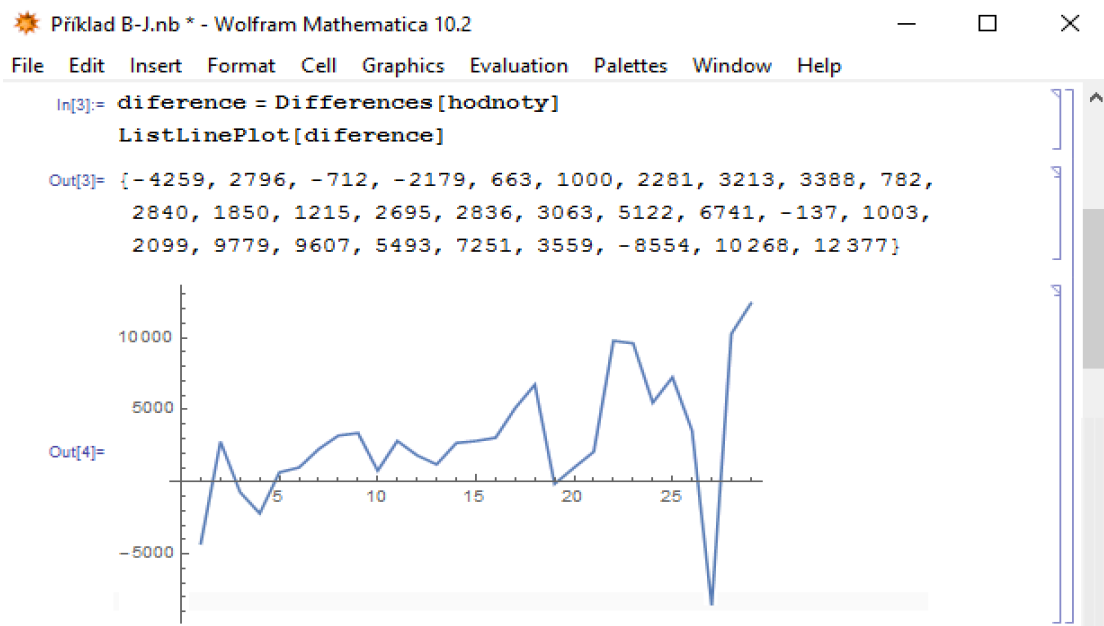
Ze všeho nejdříve je nutné provést grafickou analýzu průběhu hodnot časové řady a určit, zda je řada stacionární (*Obrázek 31*). Časová složka bude v tomto příkladu opět nahrazena pořadím hodnot.



Obrázek 31 Průběh hodnot časové řady

Zdroj: vlastní zpracování

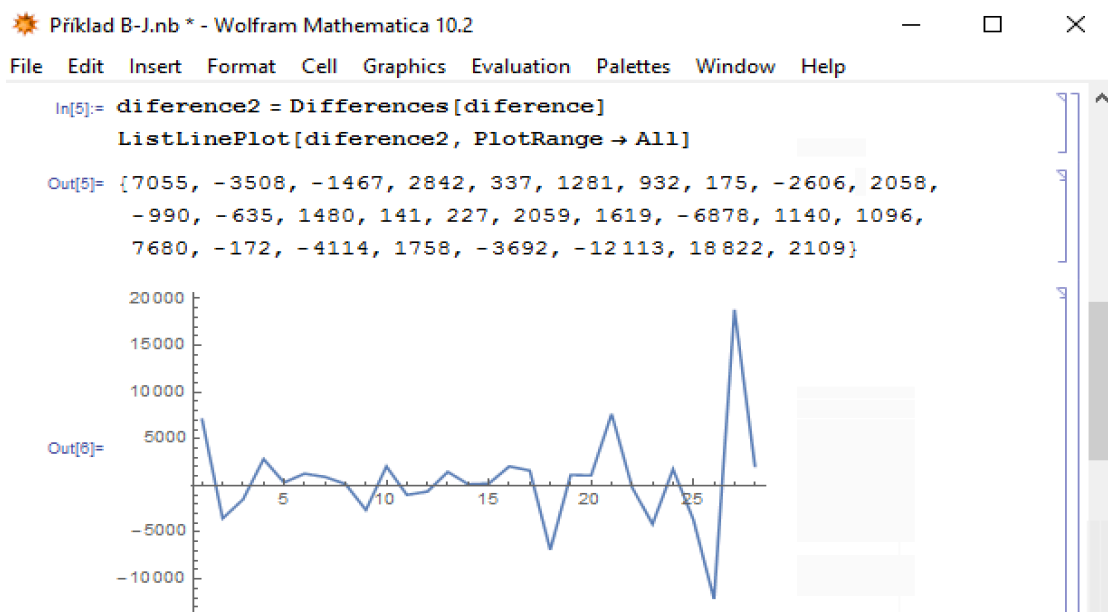
Hodnoty vykazují nestacionární průběh, proto bude dalším krokem jeho stacionarizace diferencováním.



Obrázek 32 Průběh hodnot prvních diferencí

Zdroj: vlastní zpracování

Graf ukazuje, že první diference časové řady (Obrázek 32) mají stále nestacionární průběh, proto bude sestaven graf i pro druhé diference (Obrázek 33).



Obrázek 33 Průběh hodnot druhých diferencí

Zdroj: vlastní zpracování

Na obrázku lze vidět, že to lze provést analogicky jednoduchým diferencováním již vypočtených diferencí. Hodnoty druhých diferencí se pohybují kolem vodorovné osy a s takovými hodnotami je možné dále pracovat. Z tohoto seznamu dat bude sestavena časová řada (Obrázek 34).

The screenshot shows a Mathematica notebook with the following content:

```

In[7]:= Length[diference2]
Out[7]= 28

In[8]:= t2 = Range[1, 28, 1];
        tm2 = TimeSeries[diference2, {t2}]

Out[9]= TimeSeries [ Time: 1 to 28
                    Data points: 28 ]

In[10]:= modell1 = TimeSeriesModelFit[tm2]

Out[10]= TimeSeriesModel [ Family: ARMA
                          Order: {1, 1} ]

In[11]:= eproc = modell1["Process"]
Out[11]= ARMAProcess [516.985, {0.129864}, {-0.767523}, 1.91089 × 107]

```

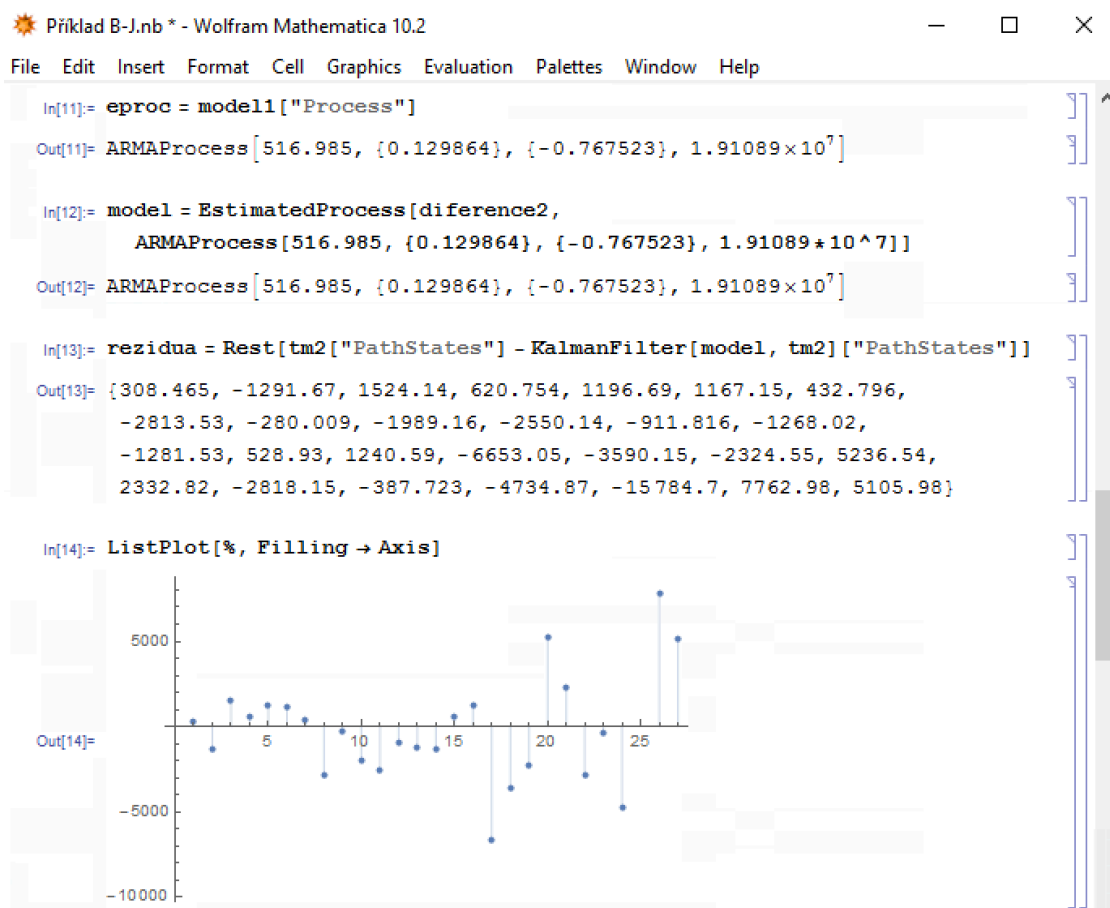
Obrázek 34 Tvorba časové řady a vyhledání vhodného modelu ARIMA

Zdroj: vlastní zpracování

Jelikož pro tvorbu časové řady byl použit příkaz *TimeSeries*, je možné modelovat řadu druhých diferencí pomocí příkazu *TimeSeriesModelFit*, který sám navrhne vhodný model. Jak je možné vidět z obrázku, software vyhodnotil, že nejvhodnějším je pro řadu model *ARMA(1,1)*, v buňce *Out[11]* jsou spočítány koeficienty tohoto procesu.

Následovat bude ověření, zda je model vhodný, to bude provedeno testem autokorelace, k tomu je ale potřeba zjistit hodnoty reziduí (Obrázek 35). Formule pro získání reziduí vyžaduje předchozí použití příkazu *EstimatedProcess*. První vstupní buňka určí konkrétní parametry modelu v návaznosti na příkaz *TimeSeriesModelFit*. Dále je použit příkaz *EstimatedProcess*, který přiřadí hodnotám seznamu *diference2* již dříve zjištěný model i s koeficienty. Je také možné nechat odhadnout koeficienty přímo příkazem *EstimatedProcess*, jejich hodnoty by ale mohly být jiné. Pro zachování jednoty tak byly přiřazeny koeficienty odhadnuté příkazem *TimeSeriesModelFit*.

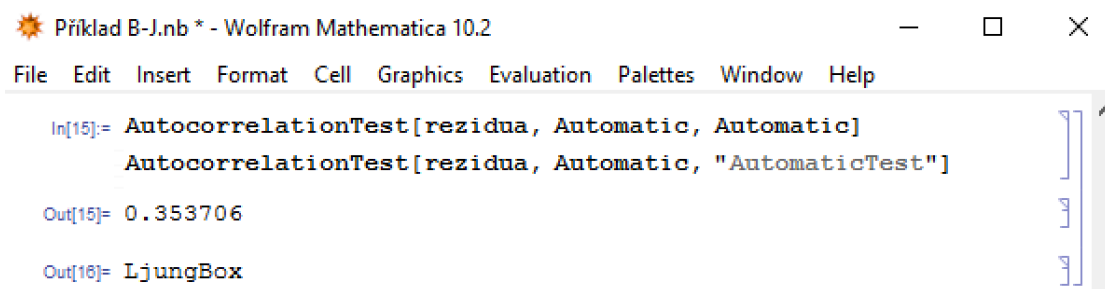
Nyní lze přistoupit k samotnému výpočtu reziduí. Pro jejich vyhodnocení slouží formule složená ze tří příkazů. Rozdíl uvnitř závorek zajišťuje nalezení rozdílu mezi hodnotami časové řady a vyhlazenými hodnotami příkazem *KalmanFilter*, kde argumenty „*Pathstates*“ určují body, ve kterých se rozdíl provádí. Příkaz *Rest* zajistí odstranění prvního členu s 0 vygenerovanou příkazem *KalmanFilter*.



Obrázek 35 Výpočet reziduí časové řady

Zdroj: vlastní zpracování

Když jsou rezidua vyhodnocena, lze otestovat model testem autokorelace (Obrázek 36). Autokorelační test lze provést prostým příkazem *AutocorrelationTest*. Ostatní parametry příkazu lze nastavit tak, aby je software Mathematica zvolil sám. Argument „*AutomaticTest*“ zajistí, že program ukáže, které kritérium pro výpočet použil. Test autokorelace v tomto případě ukazuje, že model byl ověřován Ljung-Boxovým kritériem a je pro časovou řadu vhodný.



Obrázek 36 Test autokorelace

Zdroj: vlastní zpracování

Další příklad se bude věnovat odhadům budoucího vývoje časových řad. Předpovědi budou tvořeny opět pro řadu výdajů na výzkum a vývoj v letech 1989-2018. Pro řadu již byl v předchozích příkladech vytvořen model, který vystihuje její průběh a byla také ověřena jeho vhodnost. Tento model je tudíž vhodný i pro odhad predikcí.

Příklad 9.2 (Prognózování časové řady)

Úkolem v tomto příkladu bude pro časovou řadu výdajů na výzkum a vývoj v letech 1989-2018 (*Příloha 2*) odhadnout 6 budoucích hodnot a vykreslit je společně s původní časovou řadou do grafu.

Řešení: Protože časová řada byla již pomocí Box-Jenkinsovy metodologie modelována v předchozím příkladu, lze využít některé již provedené výpočty. Software Mathematica vyhodnotil, že pro řadu druhých diferencí je vhodným modelem proces *ARMA*. Tento proces není autokorelován je proto vhodný také pro predikce. Zjištěné informace poslouží jako základ pro odhady nových hodnot a tento příklad bude přímo navazovat na příklad předchozí (*viz Příklad 9.1*).

Lze tedy přistoupit přímo k předpovídání nových hodnot (*viz kapitola 4.3*) druhých diferencí (*Obrázek 37*). Příkaz *TimeSeriesForecast* provede odhad 4 následujících druhých diferencí podle zkonstruovaného modelu *ARMA*. Počet odhadovaných hodnot je o dvě menší než je zadáno právě proto, že se jedná o druhé diference. Při jejich zpětném načítání počet nových hodnot původní časové řady nabyde právě o 2 chybějící. Nové hodnoty byly vyčísleny argumentem "*Path*", který se váže k příkazu *TimeSeriesForecast*, a jsou zobrazeny v obrázku ve výstupní buňce *Out[19]*.

```

B-J předpovědi 1.nb * - Wolfram Mathematica 10.2
File Edit Insert Format Cell Graphics Evaluation Palettes Window Help
In[18]:= předpověď = TimeSeriesForecast[model, difference2, {4}]
Out[18]= TemporalData [
  Time: 28 to 31
  Data points: 4 Paths: 1
]
In[19]:= předpověď["Path"]
Out[19]= {{28, -3127.86}, {29, 110.789}, {30, 531.373}, {31, 585.991}}

```

Obrázek 37 Odhadnuté druhé difference

Zdroj: vlastní zpracování

Následovat bude výpočet odhadnutých hodnot časové řady (Obrázek 38). Obrázek ukazuje výpočet odhadnutých hodnot prvních diferencí i přímo časové řady, nakonec jsou původní i nové hodnoty vykresleny do grafu. Nové hodnoty časové řady jsou získávány použitím příkazu *Accumulate*. Poprvé je pro ukázkou použit pro získání odhadnutých prvních diferencí, ty však již dále nejsou zapotřebí, proto nebyl ani vypsány. Podruhé akumuluje poslední známou hodnotu časové řady, poslední známou první diferenci a odhadnuté druhé difference. Výsledky tohoto výpočtu jsou uvedeny v buňce *Out[21]*. Odhadnuté hodnoty jsou pak připojeny k ostatním, již dříve známým hodnotám příkazem *Join*. Nakonec jsou znázorněny grafem. I zde bude časová složka nahrazena pořadím hodnot.

```

B-J předpovědi 1.nb * - Wolfram Mathematica 10.2
File Edit Insert Format Cell Graphics Evaluation Palettes Window Help
In[20]:= novédifference =
  Accumulate[{12377, -3127.86, 110.789, 531.373, 585.991}];
novéhodnoty =
  Accumulate[{102754, 12377, -3127.86, 110.789, 531.373, 585.991}]
Out[21]= {102754, 115131, 112003., 112114., 112645., 113231.}
In[22]:= čřpředpověď = Join[hodnoty, novéhodnoty];
In[23]:= ListLinePlot[čřpředpověď]
Out[23]=

```

Obrázek 38 Předpovězené hodnoty časové řady

Zdroj: vlastní zpracování

Příklad ukázal, jak je možné pomocí diferencí předpovědět hodnoty pro časovou řadu, jak je lze vyčíslit a i s původními hodnotami zakreslit do grafu.

Závěr

Tématem této bakalářské práce je využití softwaru Mathematica při analýze časových řad. Cílem bakalářské práce bylo zjistit, které příkazy softwaru Mathematica jsou využitelné pro vyhodnocování časových řad, popsat je a demonstrovat na příkladech s konkrétními časovými řadami.

Jednotlivé příkazy jsou uvedeny už v teoretické části této práce přímo, a to v těch kapitolách, ve kterých se nachází problematika, ke které se příkazy vztahují. Závěrečná část každé kapitoly označená jako „SM – ukázka syntaxe příkazů“ obsahuje podrobný popis syntaxe příkazů. V praktické části je pak na příkladech demonstrováno jejich použití. Software Mathematica není přímo statistický program, proto bylo zapotřebí v praktické části práce pro výpočet některých ukazatelů sestavovat formule složené z více příkazů. Tato možnost však ukazuje flexibilitu programu Mathematica.

Analýza časových řad byla zaměřena na jejich základní charakteristiky a pro jejich modelování na metodu dekompozice a Boxovu-Jenkinsovu metodologii. Pro tyto dva přístupy k modelování časových řad byly nejprve popsány jejich hlavní principy a později postup při vytváření modelu. Existuje řada dalších metod modelování, kritérií pro ověřování a ukazatelů pro analýzu časových řad, rozsah práce však nedovolil zacházet do větší hloubky.

Praktická část bakalářské práce se skládá z jednotlivých příkladů. Ty byly nejprve zaměřené na generování, import a export dat. Pak byla prováděna analýza časových řad. Většina výpočtů ukazatelů a modelování byla prováděna na časové řadě výdajů na výzkum a vývoj v České republice v letech 1989-2018. Zde byly zkoumány základní charakteristiky časové řady a konstruovány modely pomocí obou popsaných metod. Byl také uveden příklad, který demonstruje, jakým způsobem lze z časové řady odstranit sezónní vlivy. Pro tyto účely byla použita čtvrtletní časová řada počtu hostů hromadných ubytovacích zařízení v České republice v letech 2012-2018.

V části modelování časových řad metodou dekompozice byla pro časovou řadu vybrána a ověřována trendová funkce. Pro porovnání měr přesnosti, indexu determinace a modifikovaného indexu determinace byl také zkonstruován další model. Výsledky ukázaly, že vhodnějším z nich byl zvolený (kvadratický) model. V Box-Jenkinsově metodologii byl model vybrán přímo softwarem Mathematica, ten testu autokorelace vyhověl, byl tedy vhodný i pro následné odhadování budoucích hodnot. Predikce byla provedena v posledním příkladu, kde bylo predikováno šest hodnot následujících za posledním známým členem časové řady.

Jelikož jsou v této bakalářské práci nashromážděny příkazy a vytvořeny vyhodnocovací formule určené k analýze a modelování časových řad, může stávající bakalářská práce posloužit jako podklad pro práci s časovými řadami podporovanou nástroji programu Mathematica. V práci je uvedeno několik příkladů, které ukazují, jak je možné v softwaru Mathematica s časovými řadami pracovat. Uvedené způsoby řešení lze univerzálně použít i pro další časové řady, je však nutné věnovat pozornost interpretaci získaných výsledků. Práce tedy může sloužit každému, kdo má zájem provádět analýzu časových řad nebo rozšířit své znalosti o využití softwaru Mathematica. Příklady přinesly také konkrétní výsledky. Přínosem by mohly být predikce výdajů na vývoj a výzkum v ČR, které byly odhadovány v posledním příkladu. Užitečným výstupem mohou být také pro provozovatele ubytovacích zařízení sezónní indexy a časová řada očištěná o sezónní vlivy. Tyto i další výstupy práce mohou sloužit např. jako podklad pro řešení různých rozhodovacích úloh.

Literatura a zdroje

- ARLT, Josef, Markéta ARLTOVÁ, a Eva RUBLÍKOVÁ. *Analýza ekonomických časových řad s příklady*. 1.vyd. Praha: Vysoká škola ekonomická Praha, 2002. 148 s. ISBN 80-245-0307-7.
- ARLT, Josef a Markéta ARLTOVÁ. *Finanční časové řady*. 1. vyd. Praha: Grada Publishing, 2003. 220 s. ISBN 80-247-0330-0.
- BOUŠKA, Martin. *MATHEMATICA - příručka s příklady pro učitele a studenty* [online]. 1. vyd. Praha, 2012. 104 s. [cit. 2020-03-14]. Dostupné z: <https://www.pslib.cz/jaromir.oscadal/PRG%20-V1/Mathematica%20-%20př%C3%ADručka3.pdf>.
- CIPRA, Tomáš. *Analýza časových řad s aplikacemi v ekonomii*. 1. vyd. Praha: SNTL, 1986. 246 s. ISBN 99-00-00157-X.
- Český statistický úřad [online]. Praha, 2020 [cit. 2020-03-14]. Dostupné z: <https://www.czso.cz/csu/czso/domov>.
- HANČLOVÁ, Jana a Lubor TVRDÝ. *Úvod do analýzy časových řad* [online]. 1. vyd. Ostrava, 2003. 34 s. [cit. 2020-03-14]. Dostupné z: https://www.fd.cvut.cz/departament/k611/PEDAGOG/VSM/7_AnalyzaCasRad.pdf.
- HINDLS, Richard a kol. *Statistika pro ekonomy*. 8. vyd. Praha: Professional Publishing, 2007. 415 s. ISBN 978-808-6946-436.
- KŘIVÝ, Ivan. *Analýza časových řad*. [online]. 1. vyd. Ostrava, 2006. 77 s. [cit. 2020-03-14]. Dostupné z: http://informatika-osu.czechian.net/files/is/ancas/ANCAS_DiV.pdf.
- LITSCHMANNOVÁ, Martina. *Úvod do analýzy časových řad* [online]. 1. vyd. Ostrava, 2010. 28 s. [cit. 2020-03-14]. Dostupné z: http://homel.vsb.cz/~lit40/SMAD/Casove_rady.pdf.
- SIEGEL, Andrew F. *Practical Business Statistics*. 6. vyd. Burlington: Elsevier, 2012. 619 s. ISBN 978-0-12-385208-3.
- WELLIN, Paul a Harry CALKINS. *M101 A First Course in Mathematica*. 1. vyd. Illinois: Wolfram Research, 2011.
- Wolfram Language & System Documentation Center. *Wolfram: Computation Meets Knowledge* [online]. Wolfram, 2020 [cit. 2020-03-14]. Dostupné z: <https://reference.wolfram.com/language/>.

- Wolfram Research - Company Background. *Wolfram: Computation Meets Knowledge* [online]. Wolfram, 2020 [cit. 2020-03-14]. Dostupné z: <https://www.wolfram.com/company/background.html>.

Seznam zkratek

ACF	reziduální autokorelační funkce
In[]	vstupní buňka
MAE	průměrná absolutní chyba
MAPE	průměrná absolutní procentuální chyba
ME	průměrná chyba
MPE	průměrná procentuální chyba
MSE	průměrná čtvercová chyba
Out[]	výstupní buňka
PACF	parciální autokorelační funkce
SI	sezónní index
SM	software Mathematica

Seznam značek

a_t	náhodná složka časové řady
$\beta_0, \beta_1, \beta_2, \gamma$	parametry trendových funkcí
C_t	cyklická složka časové řady
δ_t	relativní přírůstek
Δy_t	absolutní přírůstek
$\bar{\Delta}$	průměrný absolutní přírůstek
k_t	koeficient růstu
\bar{k}	průměrný koeficient růstu
R^2	index determinace
R_M^2	modifikovaný index determinace
ρ_i	koeficienty autokorelace
S_t	sezónní složka časové řady
T_t	trendová složka časové řady
y_t	hodnoty časové řady
\bar{y}	prostý, chronologický a vážený chronologický aritmetický průměr
\hat{y}_t	odhadnuté hodnoty časové řady
Y_t	systematická složka časové řady

Seznam tabulek

Tabulka 1 Volba trendu na základě hodnot 1. a 2. diferencí.....	27
---	----

Seznam obrázků

Obrázek 1 Vstupní okno do programu	12
Obrázek 2 Ukázka výpočtu	13
Obrázek 3 Bodový graf	18
Obrázek 4 Spojnicový graf	19
Obrázek 5 Krabičkový graf	19
Obrázek 6 ACF a PACF modelu MA(1).....	38
Obrázek 7 ACF a PACF modelu AR(1).....	39
Obrázek 8 Tvary ACF a PACF pro modely AR, MA a ARMA.....	40
Obrázek 9 Generování dat z množiny reálných čísel.....	44
Obrázek 10 Generování dat z množiny celých čísel a jejich export.....	45
Obrázek 11 Import hodnot ze souboru.....	46
Obrázek 12 Prostý aritmetický průměr intervalové časové řady	49
Obrázek 13 Aritmetický chronologický průměr okamžikové časové řady	50
Obrázek 14 Formule pro absolutní a relativní přírůstky a koeficienty růstu.....	51
Obrázek 15 Výsledky absolutních a relativních přírůstků a koeficientů růstu.....	52
Obrázek 16 Průměrný absolutní přírůstek a průměrný koeficient růstu	53
Obrázek 17 Vývoj hodnot časové řady	54
Obrázek 18 Průběh prvních diferencí časové řady	55
Obrázek 19 Průběh druhých diferencí časových řad	56
Obrázek 20 Kvadratický trend.....	56
Obrázek 21 Lineární trend.....	57
Obrázek 22 Rezidua kvadratického trendu.....	58
Obrázek 23 Rezidua lineárního trendu.....	58
Obrázek 24 Míry přesnosti kvadratického a lineárního trendu	59
Obrázek 25 Index a modifikovaný index determinace pro trendové funkce.....	60
Obrázek 26 Vývoj hodnot časové řady	61
Obrázek 27 Porovnání hodnot časové řady a graf trendové a cyklické složky	62
Obrázek 28 Výpočet sezónní a náhodné složky a sezónního indexu.....	63
Obrázek 29 Hodnoty očištěné od sezónních vlivů.....	64
Obrázek 30 Porovnání hodnot časové řady a hodnot očištěných od sezónních vlivů	65
Obrázek 31 Průběh hodnot časové řady	66
Obrázek 32 Průběh hodnot prvních diferencí.....	67

Obrázek 33 Průběh hodnot druhých diferencí	67
Obrázek 34 Tvorba časové řady a vyhledání vhodného modelu ARIMA	68
Obrázek 35 Výpočet reziduí časové řady	69
Obrázek 36 Test autokorelace	70
Obrázek 37 Odhadnuté druhé diference	71
Obrázek 38 Předpovězené hodnoty časové řady	71

Seznam příloh

Příloha 1 Počet registrovaných trestných činů v letech 2000-2018 v okrese Kladno.....	83
Příloha 2 Výdaje na výzkum a vývoj v ČR v letech 1989-2018	84
Příloha 3 Počet subjektů provádějících výzkum a vývoj v ČR v letech 1995-2018.....	85
Příloha 4 Počet hostů ubytovacích zařízení v ČR v letech 2012-2018 (čtvrtletně)	86

Přílohy

Příloha 1 Počet registrovaných trestných činů v letech 2000-2018 v okrese Kladno

Rok	Hodnota
2000	5 541
2001	4 923
2002	4 916
2003	5 034
2004	4 859
2005	5 024
2006	5 537
2007	5 624
2008	4 610
2009	4 237
2010	3 878
2011	3 754
2012	3 677
2013	3 934
2014	3 360
2015	3 300
2016	3 388
2017	3 055
2018	2 737

Příloha 2 Výdaje na výzkum a vývoj v ČR v letech 1989-2018

Rok	Hodnota (mil. Kč)
1989	16 674
1990	12 415
1991	15 211
1992	14 499
1993	12 320
1994	12 983
1995	13 983
1996	16 264
1997	19 477
1998	22 865
1999	23 647
2000	26 487
2001	28 337
2002	29 552
2003	32 247
2004	35 083
2005	38 146
2006	43 268
2007	50 009
2008	49 872
2009	50 875
2010	52 974
2011	62 753
2012	72 360
2013	77 853
2014	85 104
2015	88 663
2016	80 109
2017	90 377
2018	102 754

Příloha 3 Počet subjektů provádějících výzkum a vývoj v ČR v letech 1995-2018

Rok	Hodnota
1995	570
1996	730
1997	807
1998	838
1999	1156
2000	1212
2001	1234
2002	1419
2003	1692
2004	1795
2005	1855
2006	1966
2007	2021
2008	2047
2009	2155
2010	2392
2011	2514
2012	2578
2013	2568
2014	2629
2015	2644
2016	2605
2017	2880
2018	2877

Příloha 4 Počet hostů ubytovacích zařízení v ČR v letech 2012-2018 (čtvrtletně)

Rok	Hodnota			
	1. čtvrtletí	2. čtvrtletí	3. čtvrtletí	4. čtvrtletí
2012	2 833 054	3 938 565	5 202 890	3 124 308
2013	2 896 839	3 934 777	5 342 897	3 233 158
2014	2 725 815	4 068 654	5 457 273	3 335 334
2015	2 979 255	4 447 245	6 181 248	3 587 802
2016	3 246 678	4 585 326	6 616 273	3 940 576
2017	3 565 691	5 228 012	6 961 294	4 245 564
2018	3 941 858	5 485 444	7 332 542	4 487 306

ANOTACE

Bibliografický údaj: Šauerová, Veronika. *Časové řady a jejich vyhodnocování pomocí softwaru Mathematica*. Olomouc 2020. Bakalářská práce. Moravská vysoká škola Olomouc. Vedoucí práce: Mgr. Jan Wossala, Ph.D.

Název práce: Časové řady a jejich vyhodnocování pomocí softwaru Mathematica

Autor: Veronika Šauerová

Ústav: Ústav informatiky a aplikované matematiky

Vedoucí práce: Mgr. Jan Wossala, Ph.D.

Počet stran: 88

Rok obhajoby: 2020

Abstrakt:

Tato bakalářská práce se zabývá využitím Softwaru Mathematica při analýze časových řad. Cílem stávající bakalářské práce je zjistit, které příkazy Softwaru Mathematica jsou využitelné pro vyhodnocování časových řad, popsat je a demonstrovat na příkladech s konkrétními časovými řadami. Práce je rozdělena do dvou částí – teoretické a praktické. První z nich obsahuje základní údaje o softwaru Mathematica, matematický rozbor charakteristik časových řad a vybrané metody pro konstrukce modelů časových řad. Analýza je zaměřena na základní charakteristiky časových řad, metodu dekompozice a Box-Jenkinsovu metodologii. Principy těchto charakteristik a metod jsou v práci popsány a demonstrovány na příkladech ve druhé části práce. Nejdůležitější prvky pro vyhodnocování vybraných ukazatelů v softwaru Mathematica jsou vyhledané příkazy a zkonstruované formule.

Klíčová slova

Box-Jenkinsova metodologie, časová řada, metoda dekompozice, software Mathematica, základní charakteristiky

Title: Time series and their solving by means of software Mathematica

Author: Veronika Šauerová

Department: Department of Computer Science and Applied Mathematics

Supervisor: Mgr. Jan Wossala, Ph.D.

Number of pages: 88

Presentation and defence: 2020

Abstract:

This bachelor thesis deals with using the Mathematica software in time series analysis. The aim of this thesis is to find out which commands of the Mathematica software are utilizable in regard to the time series evaluation, to describe them and to demonstrate them on examples with specific time series. The thesis is divided into two parts - theoretical and practical. The first one contains basic data about Mathematica Software, mathematical analysis of time series characteristics and selected methods for construction of time series models. The analysis focuses on basic characteristics of time series, decomposition method and Box-Jenkins method. Principles of these characteristics and methods are described and demonstrated on examples in the second part of this thesis. The most important elements for evaluating selected indicators in the Mathematica software are given commands and constructed formulas.

Keywords:

Box-Jenkins method, time series, decomposition method, software Mathematica, main characteristics