

Univerzita Hradec Králové
Přírodovědecká fakulta
Katedra matematiky

Odhady regresní funkce vývoje časových řad

Bakalářská práce

Autor: Petra Šupolová
Studijní program: B1103/Aplikovaná matematika
Studijní obor: Finanční a pojistná matematika
Vedoucí práce: Mgr. Jitka Kühnová, Ph.D.

Prohlášení:

Prohlašuji, že jsem bakalářskou práci vypracovala samostatně a že jsem v seznamu použité literatury uvedla všechny prameny, z kterých jsem vycházela.

V Hradci Králové dne

Petra Šupolová

Anotace

Obsahem této bakalářské práce je popis parametrických a neparametrických metod pro odhad regresní funkce vývoje časových řad. Následně jsou tyto metody aplikovány na reálná čísla z finančních trhů. Závěrem práce je zhodnocení jednotlivých metod.

Klíčová slova

Časové řady, regresní model, metoda nejmenších čtverců, klouzavé průměry, jádrové odhady.

Annotation

The content of this bachelor thesis is to describe parametric and nonparametric methods to estimate the regression function of the time series development. Afterwards, these methods are to be applied to the real data from financial markets and, lastly, evaluated individually.

Keywords

Time series, regression model, least squares method, moving averages, kernel estimations.

Obsah

Úvod	7
1 Časové řady	8
1.1 Pojem a druhy časových řad	8
1.2 Elementární charakteristiky časových řad	9
1.3 Analýza časové řady	11
1.4 Trendová složka	13
1.5 Sezónní složka	14
1.6 Cyklická složka	15
1.7 Náhodná složka	15
2 Metody odhadu trendu u časových řad	17
2.1 Regresní model	17
2.1.1 Lineární regresní model	18
2.1.2 Metoda nejmenších čtverců	18
2.2 Klouzavé průměry	21
2.2.1 Prosté klouzavé průměry	22
2.2.2 Vážené klouzavé průměry	23
2.2.3 Centrované klouzavé průměry	23
2.3 Jádrové odhady	23
3 Rozbor časových řad	26
3.1 Analýza dat o počtu domácností s vlastním osobním počítačem	26
3.1.1 Zkoumání elementárních charakteristik	27
3.1.2 Proložení časové řady regresními polynomy	28
3.1.3 Vyhlazení časové řady pomocí klouzavých průměrů	30
3.1.4 Proložení vyhlazených hodnot regresními polynomy	31
3.1.5 Využití jádrových odhadů	32
3.1.6 Shrnutí	34
3.2 Analýza dat o počtu uchazečů o zaměstnání	36
3.2.1 Zkoumání elementárních charakteristik	36
3.2.2 Proložení časové řady regresními polynomy	38
3.2.3 Vyhlazení časové řady pomocí klouzavých průměrů	38

3.2.4	Proložení vyhlazených hodnot regresními polynomy	39
3.2.5	Využití jádrových odhadů	40
3.2.6	Shrnutí	41
3.3	Analýza dat o kurzu měny	43
3.3.1	Zkoumání elementárních charakteristik	43
3.3.2	Proložení časové řady regresním polynomem	44
3.3.3	Vyhlazení časové řady pomocí klouzavých průměrů	45
3.3.4	Proložení vyhlazených hodnot regresním polynomem	46
3.3.5	Využití jádrových odhadů	47
3.3.6	Shrnutí	48
3.4	Tabulka vypočítaných hodnot	49
	Závěr	50
	Seznam použité literatury	54

Úvod

Časové řady se objevují v mnoha situacích. Jsou tvořeny opakovaným shromažďováním informací. Již z této věty lze usoudit, že časové řady je možné hledat ve statistických úřadech, které zaznamenávají data během několika let v různých oblastech (např. finanční hospodaření, inflace, nezaměstnanost, sčítání lidu atd.). Data nejsou zaznamenávána pouze na statistických úřadech, ale například i ve firmách nebo ústavech, které zaznamenávají své výkony.

V první kapitole této práce je definován pojem časové řady. Je vysvětleno, jak je možné je řadit, pomocí jakých charakteristik si o nich lze vytvořit základní představu, jaký mají tvar a z čeho se skládají.

V druhé kapitole jsou řešeny metody odhadů trendů u časové řady. Pomocí trendu se určuje základní tvar časové řady. Jednotlivé metody se dělí na parametrické a neparametrické. U obou těchto skupin se určí, co se do nich dá zařadit. Do skupiny parametrických metod se řadí lineární regresní model a metoda nejmenších čtverců. Do neparametrické skupiny se zařadí metoda klouzavých průměrů a jádrové odhady.

V poslední kapitole jsou všechny tyto poznatky prakticky vyzkoušeny na datech získaných z Českého statistického úřadu.

V první části této práce byly především získány informace z publikací Statistické metody v tržním hospodářství [10] a Ekonomické časové řady [2]. Informace o jádrových odhadech jsou získány z internetových stránek Institutu biostatiky a analýz Masarykovy Univerzity [8].

Pro rozbor časových řad byla data získána na stránkách Českého statistického úřadu [7] a na stránkách Kurzy.cz [6], které zveřejňují aktuální směnné kurzy ČNB.

Kapitola 1

Časové řady

1.1 Pojem a druhy časových řad

Časové řady nazývané také jako ekonomické časové řady jsou posloupnosti srovnatelných pozorování (dat) vymezeného ekonomického ukazatele. Časové řady závisí na čase, ve kterém byly získány. Ekonomické ukazatele jsou v oblasti politické, ekonomické nebo například v meteorologii, zemědělství, medicíně atd.

Hodnoty v časové řadě jsou uspořádány podle doby získání v čase od minulosti do přítomnosti a jsou jednoznačně a shodně měřeny na identických objektech, které jsou předmětem zkoumání. Základním cílem pozorování hodnot je odhadnutí jejich vývoje do budoucna. Hodnoty získané měřením se zapisují následujícím způsobem

$$y_1, y_2, \dots, y_n$$

nebo

$$y_t, \quad t = 1, 2, \dots, n,$$

kde y udává dané hodnoty a t je čas, ve kterém byly naměřeny od 1 do n . V publikacích je možné nalézt kromě zápisu y_t také označení $y(t)$.

Ekonomické časové řady se rozlišují podle charakteru dat na **intervalové časové řady** a **okamžikové časové řady**. U intervalových časových řad naměřené hodnoty závisí na délce časového intervalu sledování (např. rok, čtvrtletí, měsíc, atd.). Oproti tomu okamžikové časové řady se vztahují k určitým časovým okamžikům a nezávisí na délce intervalu sledování. Příkladem intervalových časových řad může být objem výroby určitého výrobku podniku za čtvrtletí nebo náklady vynaložené na měsíční výrobu. Okamžikovou časovou řadou se pak rozumí například počet zaměstnanců podniku k určitému datu nebo množství peněžních prostředků v pokladně.

Další možností dělení časových řad je rozdělení podle délky intervalu sledování. **Dlouhodobé časové řady** jsou sledovány v delších časových obdobích (např. v letech), naopak **krátkodobé časové řady** jsou pozorovány v kratších

časových obdobích než jeden rok (např. ve čtvrtletích, v měsících nebo v týdnech). S délkou intervalu sledování souvisí i tvar časových řad. U dlouhodobých pozorování bude tvar časové řady hladší oproti řadám s krátkodobým pozorováním, kde tvar bude spíše schodovitý. Speciálním typem krátkodobých časových řad jsou **vysokofrekvenční časové řady**, které jsou sledované v úsecích kratších, než je jeden týden. [3]

Do vysokofrekvenčních časových řad se zařadí tzv. **finanční časové řady**. Tyto řady jsou tvořeny z cen a jejich vývoje na finančních trzích. Hodnoty jsou získávány s vysokou frekvencí, nejčastěji to bývá s denní frekvencí. Jako příklad se uvádí směnné kurzy. U finančních časových řad lze pozorovat trend a cykličnost, o kterých se bude podrobněji hovořit dále. [3]

Hodnoty získané pozorováním se také rozdělují podle druhu na **absolutní ukazatele** a **odvozené ukazatele**. Absolutní ukazatelé jsou jednotlivé naměřené hodnoty a odvození ukazatelé jsou součtové či podílové hodnoty. A dále lze podle způsobu vyjádření dělit hodnoty na hodnoty měřené v **naturálních jednotkách** nebo **peněžních jednotkách**.

1.2 Elementární charakteristiky časových řad

Pro získání základní představy o chování časových řad se používají elementární charakteristiky časových řad. Jedná se o rychlé zhodnocení dat.

K nejjednodušším charakteristikám časových řad patří výpočet jejich průměru. Pro pozorování dat okamžikových se získá průměr sledované časové řady pomocí výpočtu chronologického průměru

$$\bar{y} = \frac{\frac{y_1}{2} + y_2 + \dots + y_{n-1} + \frac{y_n}{2}}{n - 1},$$

kde y_1, y_2, \dots, y_n jsou hodnoty jednotlivých pozorování a n je počet pozorování. Jsou-li mezi okamžikovými pozorováními stejně dlouhé časové rozestupy, pak se hovoří o **prostém chronologickém průměru**. V případě různě dlouhých intervalů se používá **vážený chronologický průměr**

$$\bar{y} = \frac{\frac{y_1+y_2}{2}(t_2 - t_1) + \frac{y_2+y_3}{2}(t_3 - t_2) + \dots + \frac{y_{n-1}+y_n}{2}(t_n - t_{n-1})}{t_n - t_{n-1}},$$

kde y_1, y_2, \dots, y_n jsou hodnoty jednotlivých pozorování, n je počet pozorování a t_1, t_2, \dots, t_n jsou jednotlivé časové okamžiky.

U intervalových časových řad také dochází k problémům při srovnávání různě dlouhých intervalů (např. srovnávání dat za měsíc leden a únor, kdy každý z těchto měsíců má různý počet dní). Proto se různě dlouhé intervaly přepočítávají na **jednotkový časový interval**. Tomuto procesu se říká **očišťování časových řad** na kalendářní dny, popřípadě na obchodní (pracovní) dny. Očišťování časových

řad na obchodní (pracovní) dny se používá především v oblasti obchodu.

Údaje očištěné na kalendářní dny získáme pomocí vzorce:

$$y_t^{(0)} = y_t \frac{\bar{p}_t}{p_t},$$

kde $y_t^{(0)}$ jsou očištěné hodnoty na kalendářní dny, údaj y_t představuje naměřenou hodnotu ukazatele v příslušném sledovaném intervalu, \bar{p}_t je průměrný počet kalendářních dní v příslušném období a p_t je počet kalendářních dní ve sledovaném období, t je čas, ve kterém byly dané hodnoty naměřeny.

Pro získání očištěných dat na obchodní (pracovní) dny se zaměňují pouze průměrný počet kalendářních dní v příslušném období za průměrný počet pracovních dní v příslušném období a počet kalendářních dní ve sledovaném období za počet pracovních dní ve sledovaném období.

Za předpokladu stejné délky časových intervalů v intervalových časových řadách nebo stejné vzdálenosti mezi okamžikem zjištění dat v okamžikových časových řadách se uvažují tyto **základní charakteristiky časových řad**:

- *Absolutní přírůstky* ukazatele mezi dvěma po sobě jdoucími obdobími

$$\Delta y_t = y_t - y_{t-1}, \quad t = 2, 3, \dots, n.$$

Výpočet udává, o kolik se změnila hodnota pozorovaného období oproti předchozímu období. Vypočtená hodnota má stejné jednotky jako původní časová řada.

- *Průměrný absolutní přírůstek* ukazatele mezi dvěma po sobě jdoucími obdobími

$$\bar{\Delta} = \frac{1}{n-1} \sum \Delta y_t = \frac{y_n - y_1}{n-1}, \quad t = 2, 3, \dots, n.$$

Tato charakteristika ukazuje, o kolik se v průměru mění hodnota časové řady za časový interval.

- *Relativní přírůstek*

$$\delta_t = \frac{\Delta y_t}{y_{t-1}} = \frac{y_t}{y_{t-1}} - 1,$$

udává, o kolik procent se změnila hodnota časové řady oproti předchozímu období.

- *Koeficient růstu*

$$k_t = \frac{y_t}{y_{t-1}}, \quad t = 2, 3, \dots, n.$$

Výpočtem se získá míra, se kterou se změnila poměřovaná hodnota časové řady. U koeficientu růstu se nezachovávají stejné veličiny, ale udávají se v procentech či desetinných číslech.

- *Průměrný koeficient růstu*

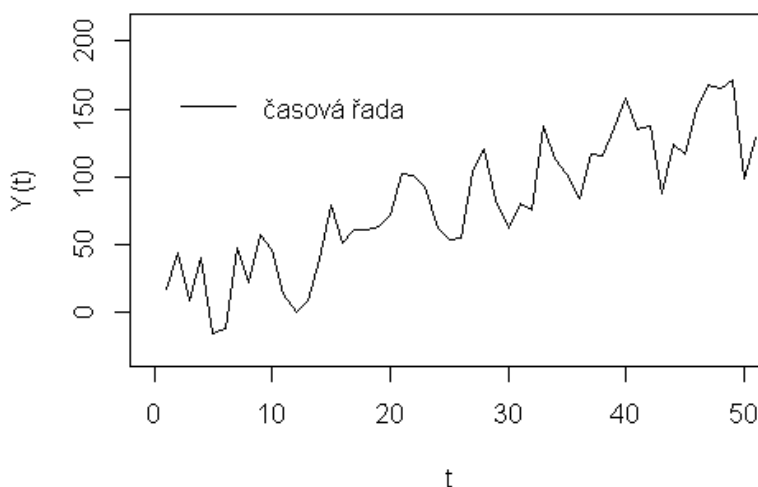
$$\bar{k} = \sqrt[n-1]{k_2 k_3 \dots k_n} = \sqrt[n-1]{\frac{y_n}{y_1}}$$

Poslední charakteristika udává, o kolik se v průměru změnil koeficient růstu časové řady. A stejně jako koeficient růstu se určuje v procentech.

1.3 Analýza časové řady

Aby se mohlo započít s analýzou časové řady, musí se nejprve zjistit, zda lze získané údaje srovnávat, tedy zda je splněno pravidlo věcné, prostorové a časové srovnatelnosti. U *věcné srovnatelnosti* se nemohou srovnávat data pořízená například před 10 lety s dnešními, které jsou technologicky dokonalejší díky nově nalezeným metodám. *Prostorovou srovnatelností* se rozumí srovnávání dat z téhož prostoru, což znamená, že se nemůže srovnávat podnik před rozdělením na menší podniky a po rozdělení, kdy se rozdělí na menší provozovny. A poslední je *časová srovnatelnost*, kdy nelze porovnávat například tržby z produkce mléka a mléčných výrobků v cenách z minulého roku s cenami současnými.

Pro získání základní představy při analýze časové řady se využívá možnosti grafického znázornění společně se základními statistickými charakteristikami. Grafické znázornění může pomoci při výpočtech nebo může být určitým druhem kontroly při výpočtech. Na obrázku 1.1 je vidět příklad grafického znázornění časové řady.



Obrázek 1.1: Grafické znázornění vzorové časové řady.

Tvar časové řady může souviset i s její **dekompozicí**. Vychází se z předpokladu, že každá časová řada je složena ze čtyř základních složek: **trendová složka**, **sezonní složka**, **cyklická složka** a **náhodná složka**. Pomocí rozkladu se identifikují pravidelná chování řad.

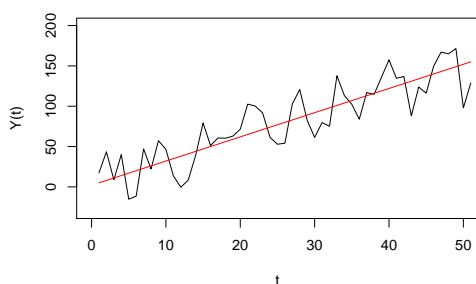
Tvar rozkladu (dekompozice) časové řady může být dvojího typu:

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

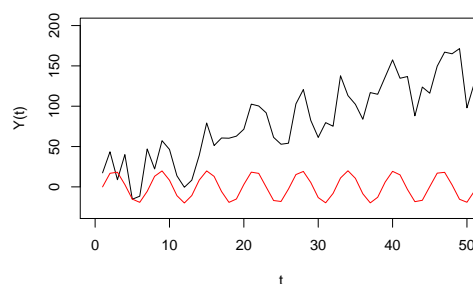
kde Y_t je daná časová řada, T_t je trendová složka, S_t znázorňuje sezónní složku, C_t složku cyklickou a ε_t je náhodná složka. Tento tvar se nazývá **aditivní tvar**, u kterého jsou hodnoty časové řady rozděleny konstantně v čase a tvar

$$Y_t = T_t \cdot S_t \cdot C_t \cdot \varepsilon_t,$$

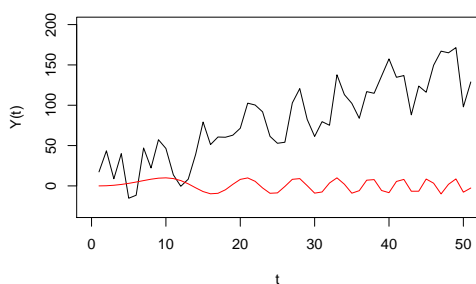
se nazývá **multiplikativní tvar** časové řady, kde hodnoty rostou v čase nebo se výrazně mění. Všechny tyto složky se však nemusí nalézt u každé časové řady hodnot.



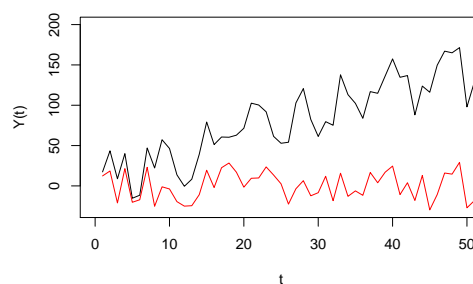
(a) Trendová složka



(b) Sezónní složka



(c) Cyklická složka



(d) Náhodná složka

Obrázek 1.2: Dekompozice časové řady $Y(t)$. Červené křivky naznačují tvary trendové, sezonní, cyklické a náhodné složky.

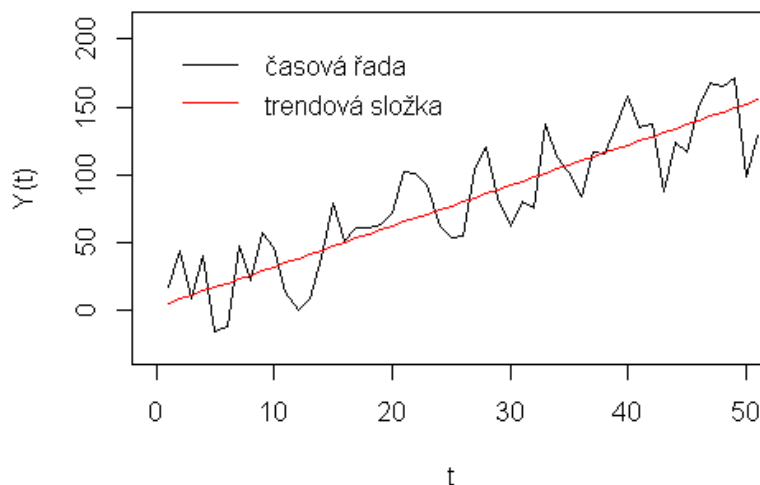
Na obrázku 1.2 je znázorněn rozklad časové řady $Y(t)$ na trendovou složku $T(t)$, sezonní složku $S(t)$, cyklickou složku $C(t)$ a náhodnou složku $N(t)$. V grafu

1.2(a) lze sledovat tvar časové řady $Y(t)$ v časovém úseku padesáti pozorování společně s tvarem trendové složky, která je znázorněna červenou křivkou. V grafu 1.2(b) je červenou čarou označen tvar sezónní složky. V grafu 1.2(c) je znázorněn tvar složky cyklické. Na posledním grafu 1.2(d) je zobrazena náhodná složka.

1.4 Trendová složka

Trendovou složkou se rozumí dlouhodobé změny v průměrném chování časových řad. Vyskytuje se, pokud se určitý faktor projevuje dlouhodobě stejným způsobem (např. určitý druh výroby výrobků). Trend může být rostoucí či klesající, mírný či strmý. V průběhu pozorování se může trend i měnit. Na obrázku 1.3 je uveden rostoucí trend u vybrané časové řady.

Časovou řadu je možno zapsat ve tvaru $Y_t = f(t) + \varepsilon_t$, $t = 1, \dots, n$, kde je $f(t)$ nazývána **trendovou funkcí**. Trendová funkce se považuje za systematickou složku a ε_t za náhodnou složku časové řady.



Obrázek 1.3: Tvar trendu u vzorové časové řady.

Rozeznávají se tyto trendové funkce:

Konstantní trend	$Y_t = \beta_0$	
Lineární trend	$Y_t = \beta_0 + \beta_1 t$	$\beta_1 \neq 0$
Kvadratický trend	$Y_t = \beta_0 + \beta_1 t + \beta_2 t^2$	$\beta_2 \neq 0$
Exponenciální trend	$Y_t = \beta_0 \beta_1^t$	$\beta_1 > 0$
Modifikovaný exponenciální trend	$Y_t = \gamma + \beta_1 \beta_2^t$	$\beta_2 > 0$
Logistický trend	$Y_t = \frac{\gamma}{1 + \beta_0 \beta_1^t}$	$\beta_1 > 0$
Gompertzova křivka	$\ln(Y_t) = \gamma + \beta_0 \beta_1^t$	$\beta_1 > 0$

Pro tyto funkce platí $t = 1, \dots, n$ a $\beta_0, \beta_1, \beta_2, \gamma \in \mathbb{R}$. Upřesňující podmínky jsou uvedeny u daných trendů.

Konstantní trend je nejjednodušší případ, kde pozorovaná data časové řady nerostou ani neklesají, pouze kolísají okolo konstanty. Tato konstanta se odhadne pomocí aritmetického průměru naměřených pozorování. V některých publikacích lze konstantní trend nalézt také pod názvem časová řada bez trendu.

U **lineárního trendu** se pozoruje vývoj hodnot s konstantním tempem růstu (lineární růst či pokles). Lineární trend se rozpozná pomocí konstantní první diference trendové funkce. Používá se k získání alespoň orientačního směru vývoje časové řady. Parametry tohoto trendu se odhadují pomocí metody nejmenších čtverců.

Kvadratický trend se určí podle druhé diference, která se bude rovnat konstantě a stejně jako u lineárního trendu se parametry získají pomocí metody nejmenších čtverců.

Metodou nejmenších čtverců se řeší i **exponenciální trendová funkce**. Metoda se aplikuje po logaritmické transformaci trendu, kterou se získá rovnice

$$\ln Y(t) = \ln \beta_0 + t \ln \beta_1.$$

Pro exponenciální trend je typické, že jeho koeficienty růstu jsou přibližně konstantní.

Pro tyto křivky platí, že nemají nenulové asymptoty a proto nejsou ohraničené a mohou růst či klesat neomezeně. Naopak pro modifikovaný exponenciální trend, logistický trend a Gompertzovu křivku lze nalézt asymptoty a proto je jejich průběh ohraničený shora nebo shora i zdola zároveň. Tyto trendy lépe odrážejí průběh časové řady.

U **modifikovaného exponenciálního trendu** se používá pro odhad parametrů trendu metoda částečných součtů, metoda dílčích průměrů nebo metoda vybraných bodů. U zmíněných metod se vždy rozdělí časová řada na tři stejně velké úseky o délce m . Pomocí součtů hodnot v těchto úsecích jsou vytvořeny tři rovnice. Řešením těchto rovnic se získají odhady parametrů.

Logistický trend, někdy nazývaný jako S-křivka podle svého typického tvaru do písmene S, lze považovat jako inverzní k modifikovanému exponenciálnímu trendu a tak lze použít stejnou výpočetní metodu pouze s hodnotami ve tvaru $\frac{1}{f(t)}$.

Gompertzova křivka patří stejně jako logistický trend do skupiny S-křivek, ale na rozdíl od zmiňovaného trendu je Gompertzova křivka asymetrická.

1.5 Sezónní složka

Sezónní složku časové řady lze určit jako pravidelné odchylky od trendové složky (promítá se v časové řadě pravidelně každé sledované období). Perioda této

složky je kratší, než celková sledovaná doba. Sezónnost lze pozorovat například při změně ročního období, době dovolených, zvyků spojených se svátky atd. U některých typů časových řad se sezónnost projevuje již přímo grafickým znázorněním časové řady (viz obrázek 1.4), u jiných to tak patrné není. Analýza sezónní složky se provádí po očištění dat od trendové funkce.



Obrázek 1.4: Tvar sezónní složky u vzorové časové řady.

1.6 Cyklická složka

Jedná se o kolísání (fáze růstu a poklesu) okolo trendu v důsledku dlouhodobého vývoje, kde je cyklus delší než jeden rok a je pozorován nepravidelně. Cykly mohou být zejména technologické, demografické či inovační. Příklad cyklické složky je zobrazen na obrázku 1.5.

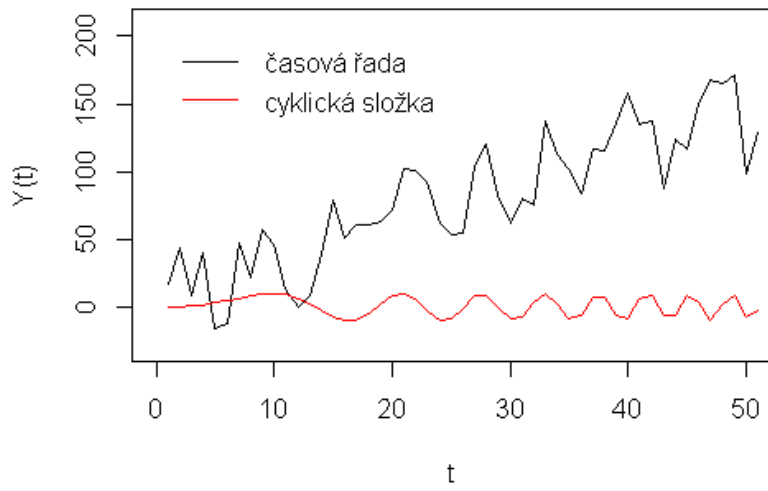
1.7 Náhodná složka

Náhodná složka, která se nazývá též složkou **reziduální** nebo **nesymetrickou**, je získána po vyloučení trendu, sezonnosti a cykličnosti z časové řady. Za náhodnou složku časové řady lze uvést chyby při měření, výpočtech či zaokrouhlování. Náhodnou složkou může být i vliv dalších neuvažovaných faktorů.

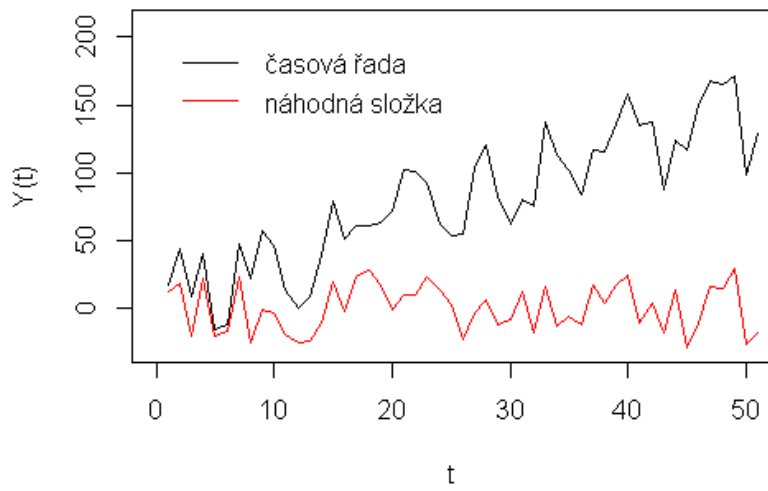
Pro náhodnou složku platí:

$$\varepsilon_t \sim N(0, \sigma^2), \quad E(\varepsilon_t) = 0, \quad D(\varepsilon_t) = \sigma^2, \quad C(\varepsilon_t, \varepsilon_{t+1}) = 0.$$

Pokud náhodná složka splňuje výše uvedené podmínky, nazývá se **bílý šum**. Reziduální složka dané časové řady je znázorněna na obrázku 1.6.



Obrázek 1.5: Tvar cyklické složky u vzorové časové řady.



Obrázek 1.6: Tvar náhodné složky u vzorové časové řady.

Kapitola 2

Metody odhadu trendu u časových řad

Nejjednodušší metodou odhadu trendu je **metoda průměrování horních a dolních bodů zvratu**. Tato metoda je založena na grafickém základu, kdy se spojují lomenými čarami horní body zvratu a dolní body zvratu (tj. lokální maxima a minima dané časové řady). Pro každý časový okamžik se následně vyznačí střed těchto vzdáleností. V některých případech však musí docházet k určité korekci hodnot (např. vynechání odlehlých pozorování nebo nevýznamné lokální body zvratu).[5]

Další metody odhadu parametrů časových řad se rozdělují do dvou skupin. První z nich je **parametrická skupina**, do které patří regresní model odhadu parametrů. Druhá skupina se nazývá **neparametrická**, do které se řadí metoda klouzavých průměrů a jádrové odhady.

2.1 Regresní model

U této metody se využívají již zmíněné trendové funkce, u kterých se dopočítávají jejich parametry. Pro výpočet parametrů trendových funkcí se nejčastěji využívá metoda nejmenších čtverců. Tuto metodu lze použít pro výpočty neznámých hodnot u funkcí konstantních, lineárních, kvadratických a dále pak pro exponenciální trendové funkce po logaritmické transformaci.

U funkcí, které nelze transformovat logaritmickou transformací ani nejsou v podobě lineární vzhledem k jejich parametrům (modifikovaný exponenciální trend, logistický trend a Gompertzova křivka) nelze metodu využít a pro jejich výpočet se používají metody jiné.

2.1.1 Lineární regresní model

Pro naměřenou matici plánů \mathbf{T} , která má rozměr $n \times k$ a naměřený vektor závislých hodnot $\mathbf{Y} = (Y_1, Y_2, \dots, Y_n)^T$, se hledá vhodný vektor parametrů $\beta = (\beta_1, \beta_2, \dots, \beta_k)$ tak, aby splňoval lineární závislost:

$$\mathbf{Y} = \mathbf{T} \cdot \beta + \epsilon,$$

kde ϵ je vektor náhodných chyb.

Obdrží se model ve tvaru:

$$Y_1 = \beta_1 t_{11} + \dots + \beta_k t_{1k} + \epsilon_1$$

:

$$Y_n = \beta_1 t_{n1} + \dots + \beta_k t_{nk} + \epsilon_n,$$

ze kterého se nalezne $\hat{\mathbf{Y}} = \mathbf{T} \cdot \hat{\beta}$, kde $\hat{\beta}$ je odhad vektoru parametrů a současně se zanedbávají náhodné chyby.

Takto zavedený model se označuje jako **lineární regresní model**.

Pokud lineární regresní model má $n > k$ a hodnota matice plánů \mathbf{T} je rovna k , pak se označuje jako **lineární regresní model plné hodnosti**. V této práci není zahrnut jiný model, než-li plné hodnosti. Jednotlivé sloupce matice plánů \mathbf{T} jsou mezi sebou nezávislé.

2.1.2 Metoda nejmenších čtverců

Je zavedený lineární regresní model ve tvaru $\mathbf{Y} = \mathbf{T} \cdot \beta + \epsilon$, pro který platí $\epsilon = \mathbf{Y} - \mathbf{T} \cdot \beta$, ϵ se nazývá reziduum.

Zavede se funkce

$$Se = \sum_{t=1}^n \epsilon_t^2 = \sum_{t=1}^n (Y(t) - \sum_{u=1}^k x_{tu} \beta_u)^2 = \|\mathbf{Y} - \mathbf{T}\beta\|^2.$$

Tato funkce se nazývá reziduální součet čtverců a nabývá svého minima v bodě $\hat{\beta} = \hat{\beta}(\mathbf{Y})$. Toto minimum funkce se hledá pro získání co nejlepšího odhadu regresního modelu.

Odhad $\hat{\beta}$ parametrů β v lineárním regresním modelu plné hodnosti, získaný metodou nejmenších čtverců, je řešením rovnice:

$$\mathbf{T}^T(\mathbf{T}\beta) = \mathbf{T}^T\mathbf{Y},$$

tedy

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

kde \mathbf{T}^T značí trasponovanou matici \mathbf{T} . [1]

Reziduální součet čtverců se získá pomocí vzorce

$$Se(\hat{\beta}) = (\mathbf{Y} - \mathbf{X}\hat{\beta})^T(\mathbf{Y} - \mathbf{X}\hat{\beta}) = \mathbf{Y}^T\mathbf{Y} - \mathbf{Y}^T\mathbf{X}\hat{\beta}.$$

Součet čtverců se vypočítá podle vzorce

$$S_T = \sum_{t=1}^n (Y_t - \bar{Y})^2 = \sum_{t=1}^n Y_t^2 - n\bar{Y}^2,$$

kde \bar{Y} je aritmetický průměr \mathbf{Y} .

Výraz

$$R^2 = 1 - \frac{Se}{S_T}$$

se nazývá **koeficient determinace**, který udává, jak těsná je regresní vzdálenost. Čím blíže je koeficient determinace roven jedné, tím je vzdálenost těsnější. [9]

Tato metoda se aplikuje na jednotlivé druhy trendových funkcí [10], [4]:

Konstantní trend $Y(t) = \beta_0, \quad t = 1, \dots, n$

Odhad b_0 parametru β_0 se zjistí takto:

$$\begin{aligned} \sum_{t=1}^n (Y(t) - \beta_0)^2 \\ \sum_{t=1}^n Y(t) - n\beta_0 &= 0 \\ b_0 = \bar{y} &= \sum_{t=1}^n \frac{Y(t)}{n}. \end{aligned}$$

Výpočtem je také získán aritmetický průměr naměřených hodnot.

Lineární trend $Y(t) = \beta_0 + \beta_1 t, \quad t = 1, \dots, n$

Parametry β_0, β_1 se dosadí do rovnice pro minimalizaci součtů čtverců chyb ε_t :

$$Se = \sum_{t=1}^n (Y(t) - \beta_0 - \beta_1 t)^2.$$

Tato rovnice je však funkcí neznámých parametrů a pro výpočet jejího minima je zapotřebí získání jejích prvních parciálních derivací podle β_0, β_1 a ty položit rovny nule:

$$\begin{aligned} 2 \sum_{t=1}^n (Y(t) - b_0 - b_1 t)(-1) &= 0 \\ 2 \sum_{t=1}^n (Y(t) - b_0 - b_1 t)(-t) &= 0 \end{aligned}$$

Po úpravě se získají dvě rovnice:

$$\sum_{t=1}^n Y(t) = nb_0 + b_1 \sum_{t=1}^n t$$

$$\sum_{t=1}^n Y(t)t = b_0 \sum_{t=1}^n t + b_1 \sum_{t=1}^n t^2,$$

ze kterých se vyjádří b_0, b_1 :

$$b_0 = \bar{y} - \frac{n+1}{2}b_1$$

$$b_1 = \frac{\sum_{t=1}^n ty_t - \frac{n+1}{2} \sum_{t=1}^n y_t}{\frac{n(n^2-1)}{12}}.$$

Ty se poté dosadí do rovnice lineárního trendu.

Kvadratický trend $Y(t) = \beta_0 + \beta_1 t + \beta_2 t^2, \quad t = 1, \dots, n$

Odhady b_0, b_1, b_2 parametrů $\beta_0, \beta_1, \beta_2$ se dosadí do podmínky metody nejmenších čtverců:

$$\sum_{t=1}^n (Y(t) - \beta_0 - \beta_1 t - \beta_2 t^2)^2.$$

Pro získání jejich odhadů parametrů, je stejně jako u lineárního trendu, nutné parciálně derivovat podle $\beta_0, \beta_1, \beta_2$:

$$2 \sum_{t=1}^n (Y(t) - b_0 - b_1 t - b_2 t^2)(-1) = 0$$

$$2 \sum_{t=1}^n (Y(t) - b_0 - b_1 t - b_2 t^2)(-t) = 0$$

$$2 \sum_{t=1}^n (Y(t) - b_0 - b_1 t - b_2 t^2)(-t^2) = 0.$$

Po úpravě rovnic se získají rovnice v tomto tvaru:

$$\sum_{t=1}^n Y(t) = nb_0 + b_1 \sum_{t=1}^n t + b_2 \sum_{t=1}^n t^2$$

$$\sum_{t=1}^n Y(t)t = b_0 \sum_{t=1}^n t + b_1 \sum_{t=1}^n t^2 + b_2 \sum_{t=1}^n t^3$$

$$\sum_{t=1}^n Y(t)t^2 = b_0 \sum_{t=1}^n t^2 + b_1 \sum_{t=1}^n t^3 + b_2 \sum_{t=1}^n t^4.$$

Z rovnic se vyjádří b_0, b_1, b_2 a dosadí se do předpisu kvadratického trendu.

Exponenciální trend $Y(t) = \beta_0 \beta_1^t, \quad t = 1, \dots, n$

Metodu nejmenších čtverců lze u tohoto trendu použít po logaritmické úpravě na tvar

$$\ln Y(t) = \ln \beta_0 + t \ln \beta_1$$

a jeho odhady parametrů se již snadno dopočítají. Pro statistické účely nemá lineární transformace dobré výsledky a proto se k výpočtu používá vážené metody nejmenších čtverců:

$$\sum_{t=1}^n v_t (Y(t) - \beta_0 \beta_1^t)^2,$$

kde v_t jsou předem zvolené váhy. Po zlogaritmování se minimalizuje tento výraz:

$$\sum_{t=1}^n v_t (\ln Y(t) - \ln \beta_0 - t \ln \beta_1)^2.$$

2.2 Klouzavé průměry

Metoda klouzavých průměrů se používá v takových časových řadách, které podléhají časovým změnám. Tato metoda spočívá v tom, že posloupnost naměřených pozorování se nahradí řadou průměrů vypočítaných z těchto dat. Název klouzavé průměry je odvozen z postupného výpočtu průměrů, které postupují vždy o jedno pozorování vpřed a poslední pozorování vypouští. Vypočtená hodnota klouzavého průměru se zaznamenává doprostřed klouzavé části. Vyrovnávají se tedy krátké úseky časové řady polynomickými funkcemi, čímž se časová řada vyhlazuje.

U metody se uvažují dva parametry: délka klouzavých průměrů a stupeň klouzavých průměrů. **Délka (řád) klouzavých průměrů** n vyjadřuje počet členů časové řady, ze kterých máme vypočítat průměry. **Stupeň klouzavých průměrů** vyjadřuje stupeň vyrovnávacího polynomu a značí se s .

Předpokládá se, že délka klouzavého průměru n je liché číslo, lze ho tedy vyjádřit jako $n = 2m + 1$, kde $m \in \mathbb{N}$. Pokud se vyrovnávají hodnoty sudé délky, získají se hodnoty mezi pozorováními, které nejsou naměřeny (v těchto případech se používají Centrované klouzavé průměry, viz 2.2.3). Čím větší délka průměru se zvolí, tím se získá větší vyrovnání časové řady. Negativní vlastnosti vyhlazování časových řad pomocí klouzavých průměrů lze nalézt na koncích časové řady, kde nelze z daných hodnot dopočítat koncové části vyhlazené časové řady. Tyto části budou mít délku m pozorování na začátku a na konci řady.

O stupni s klouzavých průměrů se rozhoduje pomocí diferencování dané řady:

$$\Delta y_t = y_t - y_{t-1}$$

$$\Delta^2 y_t = \Delta y_t - \Delta y_{t-1} = y_t - 2y_{t-1} + y_{t-2}$$

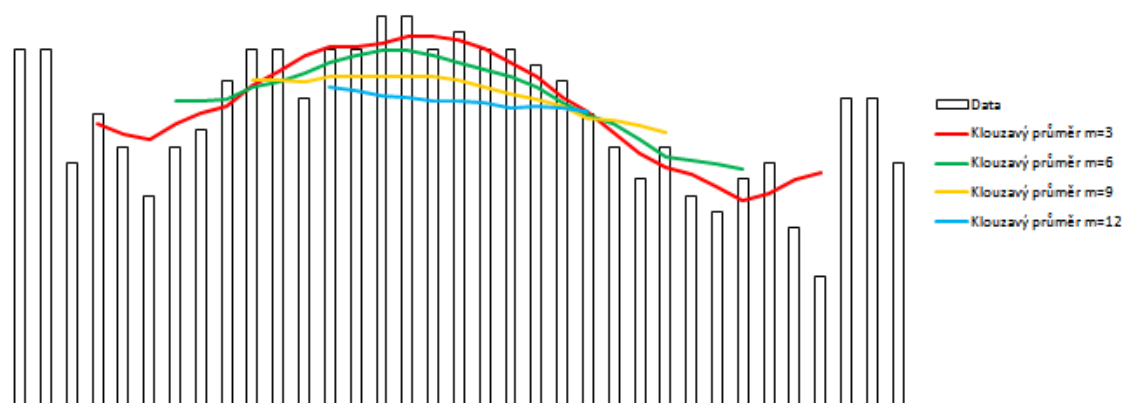
...

$$\Delta^k y_t = y_t - \binom{k}{1} y_{t-1} + \binom{k}{2} y_{t-2} - \dots + (-1)^k y_{t-k},$$

kde Δ značí první diferenci, Δ^2 druhou diferenci, atd., pak se získá vzorec:

$$V_k = \frac{\sum_{1+k}^n (\Delta^k y_t)^2}{\binom{2k}{k} (n-k)}$$

Hodnoty V_1, V_2, \dots se počítají do té doby, dokud nezačnou konvergovat k některé konstantě. U hodnot V_{r+1}, V_{r+2}, \dots , které se již blíží ke konstantě, se doporučuje zvolit klouzavé průměry stupně s . [4]



Obrázek 2.1: Ukázka klouzavých průměrů.

Na obrázku 2.1 je zobrazen výpočet klouzavých průměrů, pokud je zvolen m délky 3, 6, 9 a 12. Z grafu lze vypořadovat zkracování vyhlazené časové řady právě o m pozorování na začátku a na konci grafu.

2.2.1 Prosté klouzavé průměry

Jednotlivé úseky časové řady se nahradí prostými aritmetickými průměry:

$$\bar{y}_t = \frac{1}{2m+1} \sum_{i=-m}^m y_{t+i} = \frac{y_{t-m} + y_{t-m+1} + \dots + y_{t+m-1} + y_{t+m}}{2m+1},$$

$$t = m+1, m+2, \dots, n-m.$$

Vypočtené průměry se zamění za prostřední hodnotu v té části řady, ve které se počítá daný průměr. Pokud má časová řada sudý počet hodnot, nepoužívá se prostý klouzavý průměr, ale průměr centrovaný viz kapitola 2.2.3.

Řád klouzavé části u tohoto typu průměru se volí nejčastěji ve vzdálenosti 3, 5, nebo 7. Pokud se jedná o číselnou řadu, která obsahuje periodu, používá se jako délku klouzavé části právě velikost dané periody.

2.2.2 Vážené klouzavé průměry

Stejně jako u prostého klouzavého průměru, se hodnoty časové řady vyrovnávají v bodě t průměry, které se doplňují vahami pro jednotlivé naměřené hodnoty. Klouzavou část představuje kvadratická nebo exponenciální funkce, která určuje velikost vah. Prokládají se krátké časové úseky řady polynomem stupně s .

Vážený klouzavý průměr má tedy tvar:

$$\bar{y}_t = \sum_{i=-m}^m W_i y_{t,i}, \quad t = m+1, \dots, n-m,$$

kde W_i jsou jednotlivé váhy,

pro které platí:

$$\sum_{i=-m}^m W_i = 1$$
$$W_i = W_{-i}$$

tzn. váhy jsou symetrické.

Princip vážených klouzavých průměrů se pak aplikuje u metody jádrových odhadů.

2.2.3 Centrované klouzavé průměry

Centrované klouzavé průměry se používají v případě, že rozsah klouzavé části je sudé číslo. Tvorba centrovaného průměru je podobná jako u průměru prostého. Za účelem získání lichého počtu členů pro výpočet prostého klouzavého průměru se vytvoří z prvního a posledního členu jejich průměr:

$$y_t = \frac{\frac{y_{t-m} + y_{t+m}}{2} + y_{t-m+1} + \dots + y_{t+m-1}}{2m} = \frac{1}{4p} (y_{t-m} + 2y_{t-m+1} + \dots + 2y_{t+m-1} + y_{t+m}).$$

2.3 Jádrové odhady

Jádrové odhady patří do skupiny neparametrických metod odhadů trendové funkce, které závisí na jádrové funkci (jádře).

Pro funkci $K(t) : \mathbb{R}^n \rightarrow \mathbb{R}$ se určí nosič funkce $\text{supp}(K)$ jako uzávěr množiny

$$\{x \in \mathbb{R}^n; K(t) \neq 0\}.$$

Pro dvě nezáporná čísla ν, k platí $0 \leq \nu < k$ a reálnou funkci K s vlastnostmi:

- $K \in \text{Lip}[-1, 1]$, tj. $|K(t) - K(y)| \leq L|t - y|$ pro $\forall t, y \in [-1, 1], L > 0$
- $\text{supp}(K) = [-1, 1] \rightarrow K = 0$ mimo interval $[-1, 1]$

- funkce K splňuje tyto momentové podmínky:

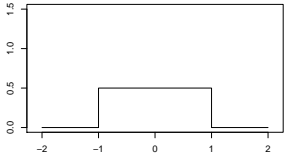
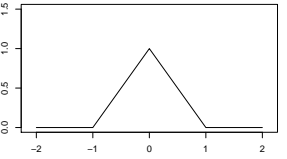
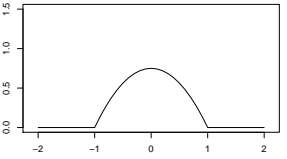
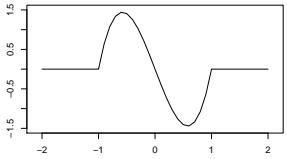
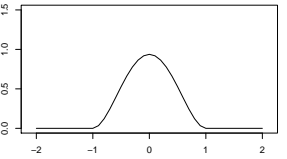
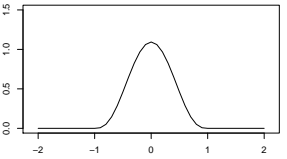
$$\int_{-1}^1 t^j K(t) dt = \begin{cases} 0 & 0 \leq j < k, j \neq \nu \\ (-1)^\nu \nu! & j = \nu \\ \beta_k \neq 0 & j = k \end{cases},$$

pak se funkce K nazývá **jádro řádu k** a **třída funkcí** se označuje $S_{\nu k}$.

Funkce $I_{[-1,1]}(t)$ je **indikátorovou funkcí** a je definována na intervalu $[-1, 1]$ takto:

$$I_{[-1,1]}(t) = \begin{cases} 1 & \text{pro } x \in [-1, 1] \\ 0 & \text{jinak} \end{cases}.$$

Tabulka 2.1: Ukázky různých jader.

<p>Obdelníkové jádro $K(t) = \frac{1}{2}I_{[-1,1]}(t)$</p> 	<p>Trojúhelníkové jádro $K(t) = 1 - t I_{[-1,1]}(t)$</p> 	<p>Epanačnikovo jádro $K(t) = \frac{3}{4}(1 - t^2)I_{[-1,1]}(t)$</p> 
<p>Jádro třídy S13 $K(t) = -\frac{15}{4}t(1 - t^2)I_{[-1,1]}(t)$</p> 	<p>Kvartické jádro $K(t) = \frac{15}{16}(1 - x^2)^2I_{[-1,1]}(t)$</p> 	<p>Hladké jádro $K(t) = \frac{35}{32}(1 - x^2)^3I_{[-1,1]}(t)$</p> 

V tabulce 2.1 se nachází ukázka jádrových funkcí společně s jejich grafy. Z předpisů funkcí lze vypořádat, že v každém předpisu funkce K se objevuje indikátorová funkce $I_{[-1,1]}(t)$ a platí pro ní výše uvedený předpis.

Obecně se definují **jádrové odhady regresní funkce m** v bodě t takto

$$\hat{m}(t, h) = \sum_{i=1}^n W_i(t, h) Y_i.$$

Funkce $W_i, i = 1, \dots, n$, která nezávisí na Y se nazývá **váhy**. Tato funkce je závislá na kladném čísle h , které se označuje jako **vyhlazovací parametr** (v jiných publikacích ho lze nalézt také pod názvem šířka vyhlazovacího okna).

Pro další výpočty se definuje tento vztah $K_h(t) = \frac{1}{h}K(\frac{t}{h})$ kde $K \in S_{0k}$ pro k je sudé.

Mezi nejznámější typy jádrových odhadů regresní funkce se řadí: Nadarayovy-Watsonovy odhady, Priestleyovy-Chaovy odhady, lokálně lineární odhady nebo Gasserovy-Müllerovy odhady. Dále se rozebírají pouze **Nadarayovy-Watsonovy odhady**.

Tyto odhady lze určit takto:

$$\hat{m}_{NW}(t, h) = \frac{\sum_{i=1}^n K_h(t - t_i) Y_i}{\sum_{i=1}^n K_h(t - t_i)}.$$

Pro bod t_0 , pro který platí $h < t_0 < 1 - h$ jsou váhy N-W odhadu zavedeny vztahem

$$W_i(t_0, h) = \frac{K_h(t_0 - t_i)}{\sum_{j=1}^n K_h(t_0 - t_j)}$$

$$\sum_{j=1}^n W_j(t_0, h) = 1.$$

Pokud se pracuje pouze na odhadu funkce m v bodech $t_i, i = 1, \dots, n$, pak pro $h \rightarrow 0$ platí:

$$\hat{m}_{NW}(t, h) = \frac{K(0)Y_i}{K(0)} = Y_i,$$

tj. při malé šířce parametru h bude odhad odpovídat více naměřeným hodnotám a

pro $h \rightarrow \infty$ platí:

$$\hat{m}_{NW}(t, h) = \frac{\sum_{j=1}^n K(0)Y_j}{\sum_{j=1}^n K(0)} = \frac{K(0) \sum_{j=1}^n Y_j}{nK(0)} = \frac{1}{n} \sum_{j=1}^n Y_j.$$

Kapitola 3

Rozbor časových řad

Tato kapitola se zabývá praktickým využitím výše uvedených vzorců na datech získaných na internetových stránkách Českého statistického úřadu. Data se zkoumají z ekonomického hlediska pomocí programu Octave, ve kterém se vypočítají jednotlivé údaje, a programu R, ve kterém se vykreslí jednotlivé grafy.

3.1 Analýza dat o počtu domácností s vlastním osobním počítačem

Jako první se prozkoumají data, která reprezentují počet domácností s vlastním osobním počítačem od roku 1989 do roku 2018. Data v tabulce jsou uvedena v procentech.

Tabulka 3.1: Počet domácností s vlastním PC.

Rok	Počet domácnosti v %	Rok	Počet domácnosti v %	Rok	Počet domácnosti v %
1989	1,8	1999	14,7	2009	54,2
1990	2,5	2000	17,9	2010	59,3
1991	3,5	2001	21,1	2011	64,8
1992	4,4	2002	24,2	2012	67,3
1993	5,1	2003	28,4	2013	68,1
1994	5,2	2004	29,2	2014	72,4
1995	6,7	2005	30,0	2015	73,1
1996	8,0	2006	35,7	2016	75,6
1997	11,0	2007	39,6	2017	76,3
1998	13,1	2008	47,7	2018	78,4

Z tabulky 3.1 je možno vyjádřit skutečnost, že daná ekonomická časová řada se řadí do dlouhodobých intervalových časových řad. Řada patří do dlouhodobých řad protože je sledována v několika letech. Do intervalových řad se řadí z

důvodu sledování během celých let. Zadané hodnoty časové řady jsou absolutní ukazatelé.

3.1.1 Zkoumání elementárních charakteristik

U hodnot se postupně určují absolutní přírůstky, průměrný absolutní přírůstek, relativní přírůstek, koeficient růstu a průměrný koeficient růstu (viz tabulka 3.2). Tyto hodnoty se nazývají odvozenými ukazateli časové řady.

Tabulka 3.2: Analýza 1 - Základní elementární charakteristiky.

t	y_t	Δy_t	k_t	δ_t	t	y_t	Δy_t	k_t	δ_t
1989	1,8	0,7	1,3889	0,3889	2004	29,2	0,8	1,0274	0,0274
1990	2,5	1,0	1,4000	0,4000	2005	30,0	5,7	1,1900	0,1900
1991	3,5	0,9	1,2571	0,2571	2006	35,7	3,9	1,1092	0,1092
1992	4,4	0,7	1,1591	0,1591	2007	39,6	8,1	1,2045	0,2045
1993	5,1	0,1	1,0196	0,0196	2008	47,7	6,5	1,1363	0,1363
1994	5,2	1,5	1,2885	0,2885	2009	54,2	5,1	1,0941	0,0941
1995	6,7	1,3	1,1940	0,1940	2010	59,3	5,5	1,0927	0,0927
1996	8,0	3,0	1,3750	0,3750	2011	64,8	2,5	1,0386	0,0386
1997	11,0	2,1	1,1909	0,1909	2012	67,3	0,8	1,0119	0,0119
1998	13,1	1,6	1,1221	0,1221	2013	68,1	4,3	1,0631	0,0631
1999	14,7	3,2	1,2177	0,2177	2014	72,4	0,7	1,0097	0,0097
2000	17,9	3,2	1,1788	0,1788	2015	73,1	2,5	1,0342	0,0342
2001	21,1	3,1	1,1469	0,1469	2016	75,6	0,7	1,0093	0,0093
2002	24,2	4,2	1,1736	0,1736	2017	76,3	2,1	1,0275	0,0275
2003	28,4	0,8	1,0282	0,0282	2018	78,4			

V prvním sloupci označeném t jsou uvedeny jednotlivé roky sledování. Ve sloupci označeném y_t jsou zaznamenány počty domácností s vlastním osobním počítačem (v %).

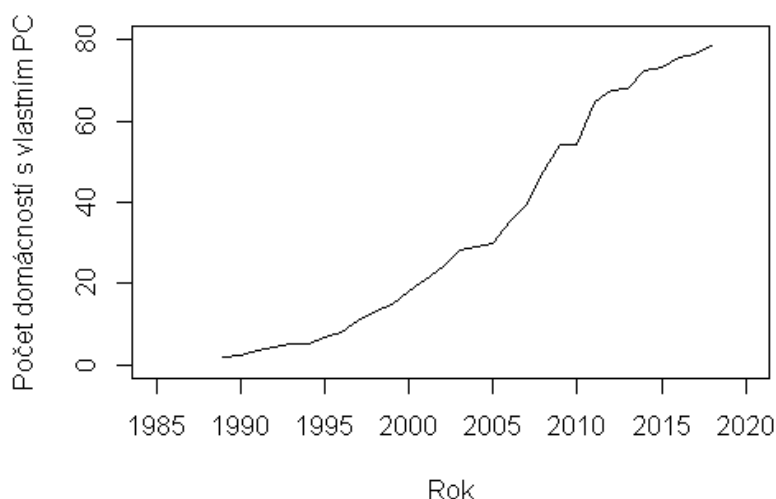
Ve třetím sloupci tabulky se pozoruje meziroční nárůst počtu domácností, které vlastní osobní počítač. Největší nárůsty v počtu počítačů, pohybující se nad hranicí 5 procent, jsou od roku 2005 do roku 2011. V předchozích letech se jednalo spíše o nižší nárůsty do 5 procent.

V dalším sloupci je sledován koeficient růstu, který v tabulce udává míru, o kolik procent se meziročně zvýšily počty domácností vlastníci osobní počítač. Toto číslo se pohybuje mezi 1,01 až 1,4.

Relativní přírůstek, udávající o kolik procent se změnila hodnota proti minulému roku, se pohybuje od 0,01 do 0,4 procentního bodu.

V průměru se hodnota časové řady za časové období od roku 1989 do roku 2018 změnila o 2.6414 %. Průměrný koeficient růstu se zvýšil o 1.1390 procentního bodu.

Už z absolutních hodnot je vidět, že každým rokem počet domácností vlastníci počítač roste. Tuto skutečnost lze pozorovat i pomocí grafického znázornění zadaných dat (viz obr. 3.1).



Obrázek 3.1: Graf počtu domácností s vlastním osobním počítačem.

3.1.2 Proložení časové řady regresními polynomy

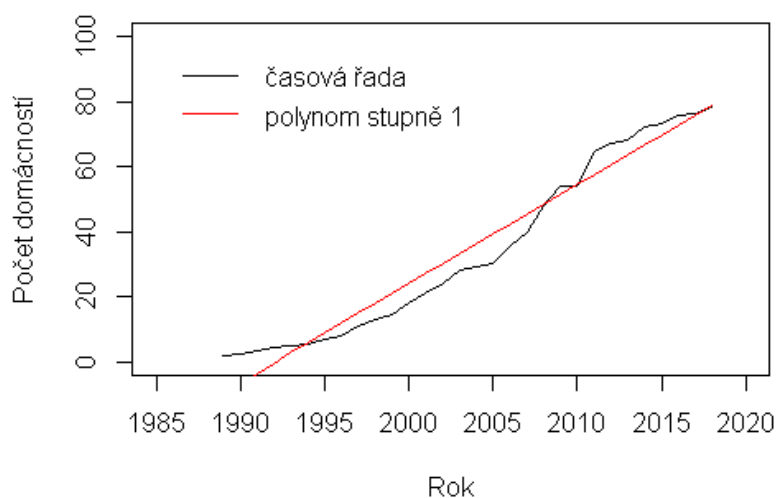
Dále se bude hovořit o prokládání hodnot v grafu polynomy s různými stupni. V rámci prokládání grafu polynomy se zkoumá i hodnota reziduálního součtu čtverců a koeficient determinace u každého stupně. Tyto hodnoty jsou zobrazeny v tabulce 3.3.

Tabulka 3.3: Analýza 1 - Výsledky proložení řady regresními polynomy.

Stupeň polynomu	1	2	3	4	5
Reziduální součet čtverců	924.48	443.25	178.32	87.528	86.345
Koeficient determinace	0.9576	0.9797	0.9918	0.9959	0.9960

Dále je z této tabulky vidět, že s rostoucím stupněm polynomu klesá hodnota reziduálního součtu čtverců. Následně lze pozorovat skutečnost, že s blížeícími se hodnotami koeficientu determinace k jedné, se proložený regresní polynom bude více blížit grafu zadaných hodnot.

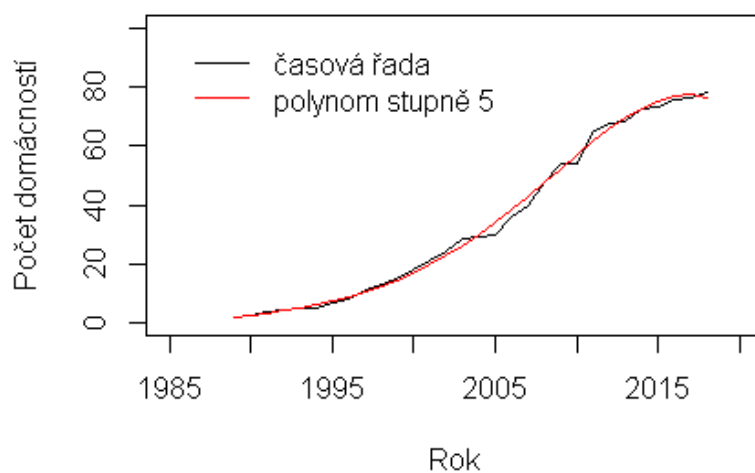
Na obrázku 3.2 je vidět prokládání hodnot polynomy stupně jedna, který ukazuje pouze směr růstu hodnot.



Obrázek 3.2: Graf domácností proložený regresním polynomem stupně 1.

Pokud se hodnoty prokládají polynomem stupně tři, hodnoty více odpovídají podobě původní řady. Na začátku a na konci časové řady se však odchyľují a mezi lety 2003 až 2012 proložený polynom příliš nezachycuje nepravidelnost původních hodnot. Z tohoto důvodu se tento regresní polynom nebude uvažovat.

Jako nejvhodnější proložení se tedy zvolí polynom stupně pět, který nejvěrněji zobrazuje tvar časové řady a koeficient determinace u tohoto stupně je nejblíže k jedné. Proložení grafu je vidět na obrázku 3.3.

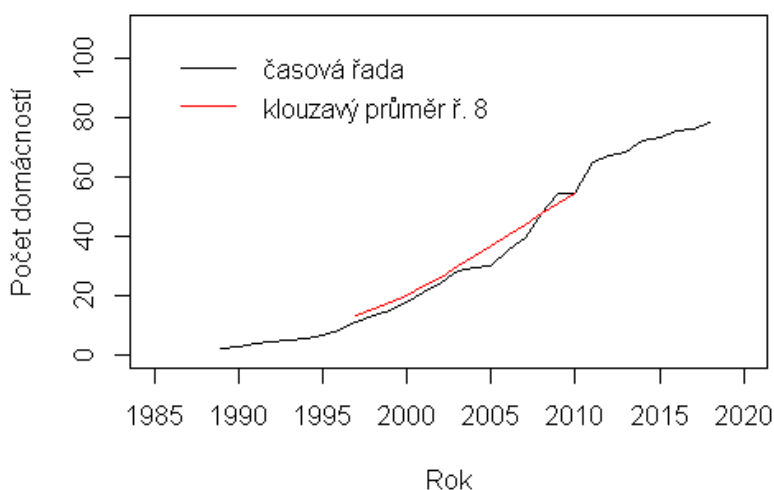


Obrázek 3.3: Graf domácností proložený regresním polynomem stupně 5.

3.1.3 Vyhlazení časové řady pomocí klouzavých průměrů

Jako další se využije metoda klouzavých průměrů. Začne se s nahrazování naměřených hodnot průměrnými hodnotami. V textu je uvedeno, že s větší délkou vyhlazování se přichází o vyhlazené hodnoty časové řady. Počet domácností s vlastním osobním počítačem je sledován po období 30 let. Nemůže se tedy uvažovat vyhlazování hodnot o délce 15 a více, z toho důvodu, že by se přicházelo o všechna data k pozorování.

Ze stejného důvodu není vhodné uvažovat vyhlazování časové řady klouzavým průměrem řádu osm a více. U těchto délek se mohou vyhlazené hodnoty již pozorovat, ale není představa o tom, jak se hodnoty vyvíjejí v koncových částech grafu (viz obr. 3.4).



Obrázek 3.4: Graf domácností proložený klouzavým průměrem délky 8.

Pomocí vzorce pro určení stupně klouzavého průměru, který je definován v kapitole 2.2), lze určit nejvhodnější stupeň klouzavého průměru.

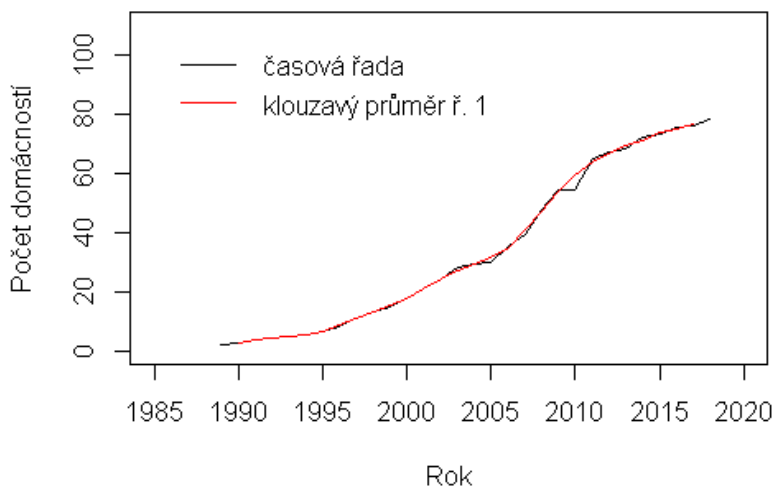
V tabulce 3.4 jsou vidět výsledky hodnoty V_k při využití různých hodnot pro k . Je uvedeno, že pokud začnou hodnoty konvergovat k některé konstantě, budou se považovat za vhodný klouzavý průměr hodnoty před konvergujícími hodnotami.

Tabulka 3.4: Analýza 1 - Vypočítané hodnoty V_k pro jednotlivé stupně vyhlazení.

k	1	2	3	4	5	6
V_k	5,527	0,699	0,634	0,632	0,632	0,622

V tomto případě je zvolen klouzavý průměr stupně jedna. Vyhlazení hodnot průměrem stupně jedna, které jsou sledovány na obrázku 3.5, přesně kopírují

hodnoty časové řady. Je zřejmé, že zejména od roku 2010 nastává vyrovnání hodnot.



Obrázek 3.5: Graf domácností proložený klouzavým průměrem stupně 1.

3.1.4 Proložení vyhlazených hodnot regresními polynomy

Hodnoty vyhlazené pomocí klouzavých průměrů se nyní považují za původní hodnoty a zkoušejí se znovu prokládat regresními polynomy. Tím, že se využijí vyhlazené hodnoty, mělo by se dosáhnout lepšího proložení.

V tabulce 3.5 je možno zkoumat hodnoty reziduálního součtu čtverců a koeficient determinace u vybraných stupňů polynomů:

Tabulka 3.5: Analýza 1 - Výsledky proložení vyhlazené řady (klouzavými průměry) regresními polynomy.

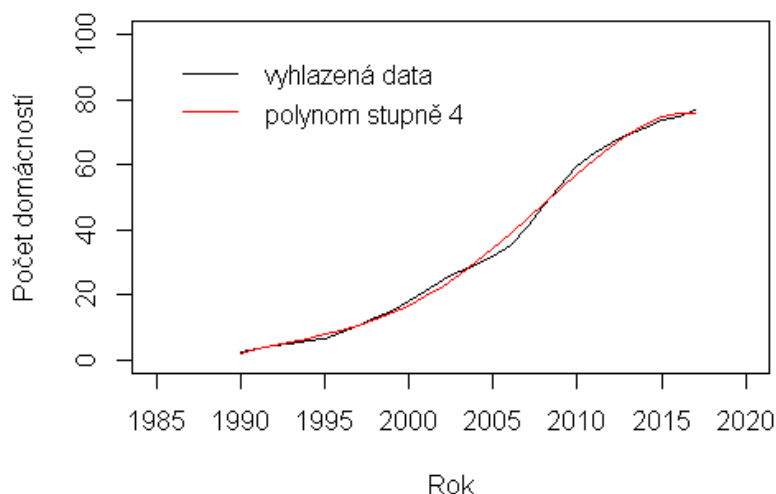
Stupeň polynomu	1	2	3	4	5
Reziduální součet čtverců	835.38	294.59	147.43	53.503	49.670
Koeficient determinace	0.9579	0.9852	0.9926	0.9973	0.9975

Pokud by se porovnávaly tabulky původních a vyhlazených hodnot, pozorovaly by se u vyrovnaných hodnot nižší výsledky reziduálního součtu čtverců. Důvody těchto rozdílů se mohou hledat ve snížení počtu pozorování o dvě hodnoty nebo vyrovnání časové řady pomocí klouzavého průměru stupně jedna.

Pokud se využije lineární polynom na vyhlazených datech, stejně jako u původních dat, bude polynom naznačovat též pouze směr růstu počtu domácností. Nelze tedy říci, že tento polynom je vhodný na proložení vyhlazených dat.

Proto jako nejvhodnější polynom, který je zvolen, má stupeň čtyři. Hodnota

koeficientu determinace u tohoto stupně polynomu není sice nejbliže k jedné (nejbliže by byl u polynomu pátého stupně), ale nejlépe kopíruje tvar původní časové řady (viz graf 3.6).



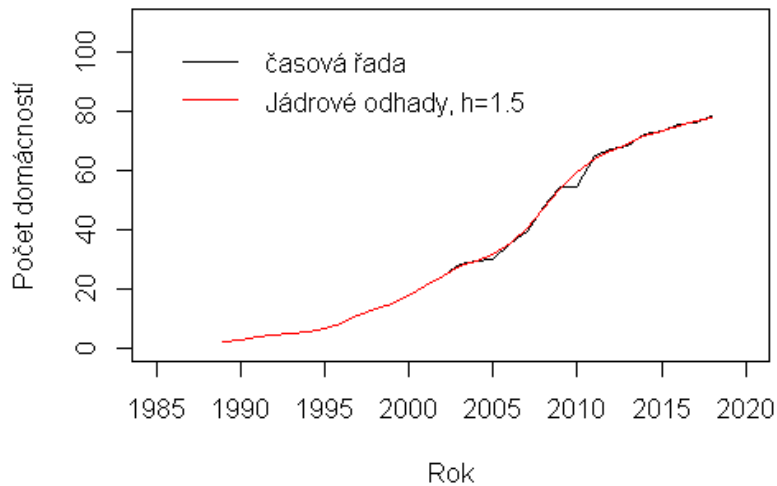
Obrázek 3.6: Graf vyhlazeného počtu domácností (klouzavým průměrem) proložený regresním polynomem stupně 4.

3.1.5 Využití jádrových odhadů

Nyní je třeba se zaměřit na další skupinu neparametrických metod a těmi jsou jádrové odhady.

Jako první je nutno si zvolit šířku vyhlazovacího okna. Při zvolení vyhlazovacího parametru h od nuly do jedné, se naměřené hodnoty při použití Nadarayových-Watsonových odhadů nijak nezmění.

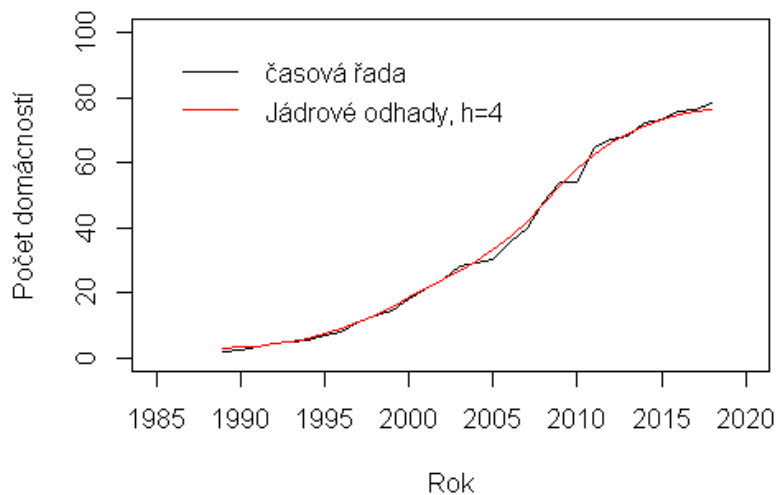
Byl zvolen vyhlazovací parametr ve výši 1.5. Vyhlazené hodnoty, pomocí jádrových odhadů, odpovídají grafu původních hodnot. Vykreslují se zároveň všechny nerovnosti jak můžeme sledovat v grafu 3.7.



Obrázek 3.7: Graf domácností proložený jádrovým odhadem s $h = 1.5$.

Je-li zvolen vyhlazovací parametr $h = 10$, je časová řada "přehlazena" a neodpovídá tvaru časové řady. Proto se tento parametr nebude používat. Musí se tedy hledat vyhlazovací parametr mezi těmito hodnotami.

Jako vhodný vyhlazovací parametr je tedy zvolen vyhlazovací parametr $h = 4$. Tento parametr kopíruje linii časové řady a není ovlivněn žádnými odchylkami (viz obr. 3.8).



Obrázek 3.8: Graf domácností proložený jádrovým odhadem s $h = 4$.

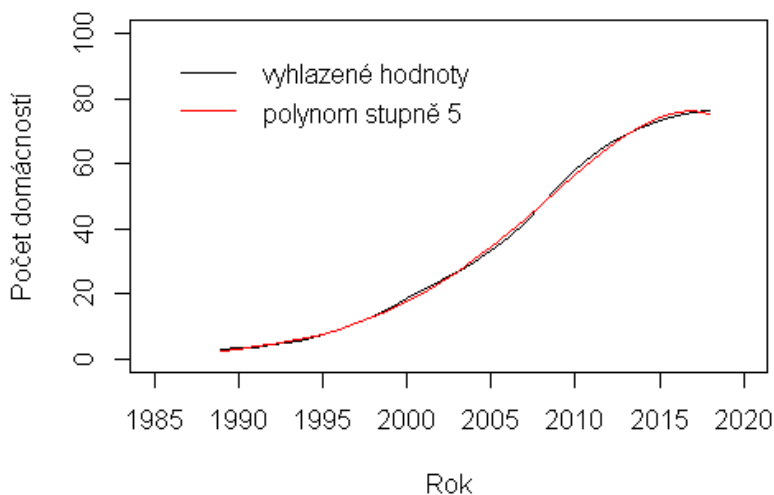
Stejně jako u klouzavých průměrů, získaný jádrový odhad s vyhlazovacím oknem čtyři, se bude prokládat regresními polynomy.

Jednotlivé hodnoty reziduálního součtu čtverců a koeficientu determinace jsou znázorněny v následující tabulce 3.6:

Tabulka 3.6: Analýza 1 - Výsledky proložení vyhlazené řady (jádrovým odhadem) regresními polynomy.

Stupeň polynomu	1	2	3	4	5
Reziduální součet čtverců	794.95	374.62	97.706	23.333	21.208
Koeficient determinace	0.9619	0.9821	0.9953	0.9989	0.9990

Už z tabulky je vidět, že jako u ostatních proložení regresními polynomy, má proložení polynomem stupně pět nejbližší hodnoty. Toto proložení je zvoleno jako nejvhodnější proložení (viz obrázek 3.9).



Obrázek 3.9: Graf vyhlazeného počtu domácností (jádrovým odhadem) proložený regresním polynomem stupně 5.

3.1.6 Shrnutí

Na závěr jsou porovnána všechna získaná data z jednotlivých prokládání. Hodnoty reziduálního součtu čtverců společně s koeficientem determinace je možné sledovat v tabulce 3.7:

Tabulka 3.7: Analýza 1 - Tabulka všech vypočítaných hodnot.

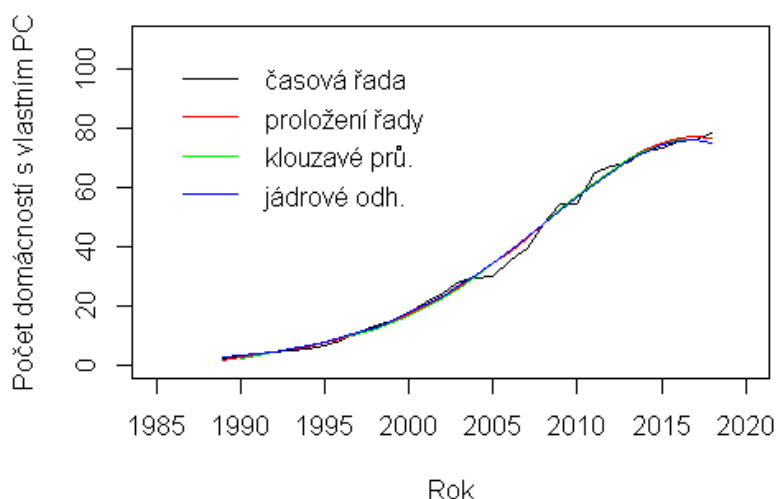
St. polynomu	Původní hodnoty		Vyhlazené klouzavými průměry		Vyhlazené jádrovými odhady	
	Se	R^2	Se	R^2	Se	R^2
1	924.48	0.9576	835.38	0.9579	794.95	0.9619
2	443.25	0.9797	295.59	0.9852	374.62	0.9821
3	178.32	0.9918	147.43	0.9926	97.906	0.9953
4	87.528	0.9959	53.503	0.9973	23.333	0.9989
5	86.345	0.9960	49.670	0.9975	21.208	0.9990

U všech tří možností reziduální součet čtverců klesá a koeficient determinace se více blíží k jedné, pokud se hodnoty budou prokládat vyššími regresními polynomy.

Z tabulky vyplývá, že výše Se u původních hodnot je proti vyhlazeným hodnotám o hodně vyšší. Nejnižší hodnoty jsou u vyhlazených hodnot pomocí jádrových odhadů s šířkou vyhlazovacího okna čtyři.

Z vypočítaných hodnot lze usoudit, že nejlepší ze všech druhů proložení je u hodnot vyhlazených pomocí jádrových odhadů s použitím polynomu stupně pět (obr. 3.10).

Za předpokladu požadavku vykreslení zvolených druhů proložení do grafu s původními hodnotami, jsou regresní polynomy téměř totožné a v grafu se překrývají.



Obrázek 3.10: Graf proložení časové řady s použitím různých metod.

3.2 Analýza dat o počtu uchazečů o zaměstnání

Jako další příklad je uvedeno množství uchazečů o zaměstnání v evidenci úřadu práce v Královéhradeckém kraji v letech 1993-2017. Data jsou vždy zjišťována k 31. prosinci daného roku.

Tabulka 3.8: Počet uchazečů o zaměstnání v Královéhradeckém kraji.

Rok	Počet uchazečů	Rok	Počet uchazečů	Rok	Počet uchazečů
1993	7306	2002	20497	2011	22185
1994	6243	2003	22146	2012	25210
1995	5674	2004	22526	2013	27678
1996	7509	2005	21989	2014	23866
1997	11236	2006	19298	2015	18574
1998	16782	2007	14499	2016	14270
1999	20783	2008	14728	2017	10483
2000	16644	2009	23373		
2001	17433	2010	24678		

Ekonomická časová řada v tabulce 3.8 je také řazena do dlouhodobých časových řad. Na rozdíl od prvního příkladu byla tato data zjišťována k určitému dni a proto se budou řadit do okamžikových časových řad. Hodnoty časové řady jsou opět absolutní ukazatelé.

3.2.1 Zkoumání elementárních charakteristik

U zadaných hodnot se budou určovat odvozené ukazatele, do kterých řadíme absolutní přírůstky, průměrné absolutní přírůstky, relativní přírůstky, koeficient růstu a průměrný koeficient růstu. Hodnoty těchto ukazatelů jsou vidět v následující tabulce 3.9.

První sloupec znázorňuje jednotlivé roky pozorování, ve druhém jsou vidět počty uchazečů.

Meziroční nárůst počtu uchazečů o zaměstnání, který lze sledovat ve třetím sloupci, se během 25 let měnil skokovitě ze záporných hodnot na kladné a opačně. Díky tomu je možné usuzovat, že graf zadaných hodnot bude proměnlivý.

Ve čtvrtém sloupci je vidět míra, o kolik procent se meziročně zvýšil počet uchazečů o zaměstnání. Tato hodnota se pohybuje mezi 0.7346 a 1.5870.

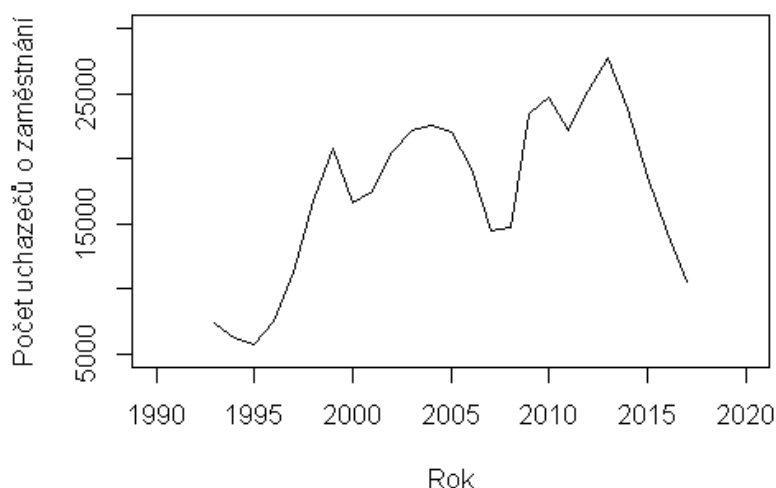
V dalším sloupci je znázorněno, o kolik procent se změnil počet uchazečů proti roku minulému. Relativní přírůstek je kolísavý stejně jako hodnota prvních derivací.

Průměrný absolutní přírůstek je ve výši 132.38 uchazečů. Průměrný koeficient růstu, který říká, o kolik se v průměru změnil koeficient růstu, má hodnotu 1.0152 procent.

Tabulka 3.9: Analýza 2 - Základní elementární charakteristiky.

t	y_t	Δy_t	k_t	δ_t	t	y_t	Δy_t	k_t	δ_t
1993	7306	-1063	0.8545	-0.1455	2006	19298	-4799	0.7513	-0.2487
1994	6243	-569	0.9089	-0.0911	2007	14499	229	1.0158	0.0158
1995	5674	1835	1.3234	0.3234	2008	14728	8645	1.5870	0.5870
1996	7509	3727	1.4963	0.4963	2009	23373	1305	1.0558	0.0558
1997	11236	5546	1.4936	0.4936	2010	24678	-2493	0.8990	-0.1010
1998	16782	4001	1.2384	0.2384	2011	22185	3025	1.1364	0.1364
1999	20783	-4139	0.8009	-0.1992	2012	25210	2468	1.0979	0.0979
2000	16644	789	1.0474	0.0474	2013	27678	-3812	0.8622	-0.1377
2001	17433	3064	1.1757	0.1758	2014	23866	-5292	0.7783	-0.2217
2002	20497	1649	1.0805	0.0805	2015	18574	-4304	0.7683	-0.2317
2003	22146	380	1.0172	0.0172	2016	14270	-3787	0.7346	-0.2654
2004	22526	-537	0.9762	-0.0238	2017	10483			
2005	21989	-2691	0.8776	-0.1224					

Za účelem získání přesné představy o struktuře dat se využívá grafického schématu (viz obr. 3.11). Jak je vidět, tyto domněnky se potvrdily a graf je opravdu kolísavý. Nejnižší počet uchazečů byl v roce 1995 a naopak nejvyšší počet v roce 2013. Z tabulky vyplývá, že během 25 let se množství lidí registrovaných na úřadu práce několikrát zvýšilo i snížilo. V současné době je pozorován velký pokles.



Obrázek 3.11: Graf počtu uchazečů o zaměstnání.

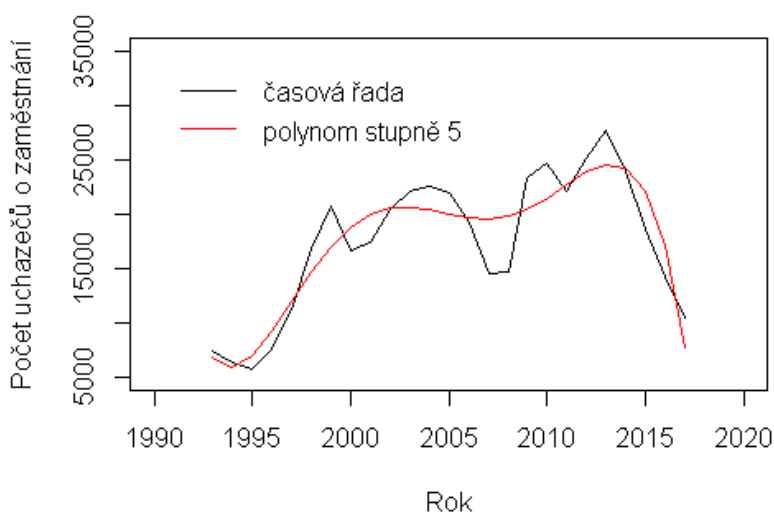
3.2.2 Proložení časové řady regresními polynomy

Graf hodnot se bude prokládat regresními polynomy s různými stupni. V tabulce 3.10 je znázorněn koeficient determinace u jednotlivých polynomů:

Tabulka 3.10: Analýza 2 - Výsledky proložení řady regresními polynomy.

Stupeň polynomu	1	2	3	4	5
Koeficient determinace	0.3082	0.6643	0.6764	0.7253	0.8421

Pokud by se časová řada prokládala polynomem vyššího řádu než pět, začnou se první sloupce matice plánů proti ostatním sloupcům zmenšovat až začnou být zanedbatelné. Tudíž je zvoleno proložení polynomem stupně pět (viz obr. 3.12).



Obrázek 3.12: Graf počtu uchazečů proložení regresním polynomem stupně 5.

3.2.3 Vyhlazení časové řady pomocí klouzavých průměrů

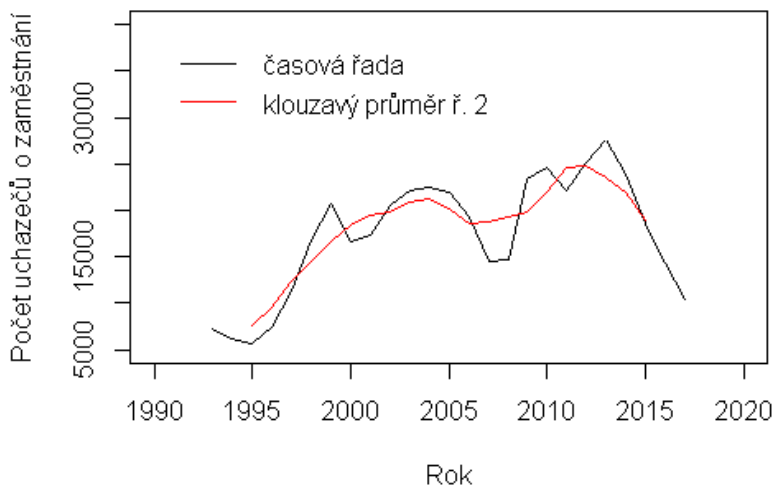
Prostřednictvím vzorce pro určení stupně klouzavého průměru (viz kapitola 2.2) se zjistí, jaký stupeň se má použít při úpravě časové řady:

Tabulka 3.11: Analýza 2 - Vypočítané hodnoty V_k pro jednotlivé stupně vyhlazení.

k	1	2	3	4	5	6
V_k	6.2374	2.6275	1.7292	1.2786	1.0072	0.8381

Z tabulky 3.11 je možné vyvodit, že vhodný stupeň vyhlazení časové řady je druhý. Na obrázku 3.13 lze vidět, jak se časová řada změní při tomto kroku.

Přišlo se o dvě pozorování na začátku časové řady a dvě pozorování na konci časové řady. Vyhladily se skoky zejména v letech 1999, 2007 a 2008.



Obrázek 3.13: Graf počtu uchazečů vyhlazený pomocí klouzavého průměru stupně 2.

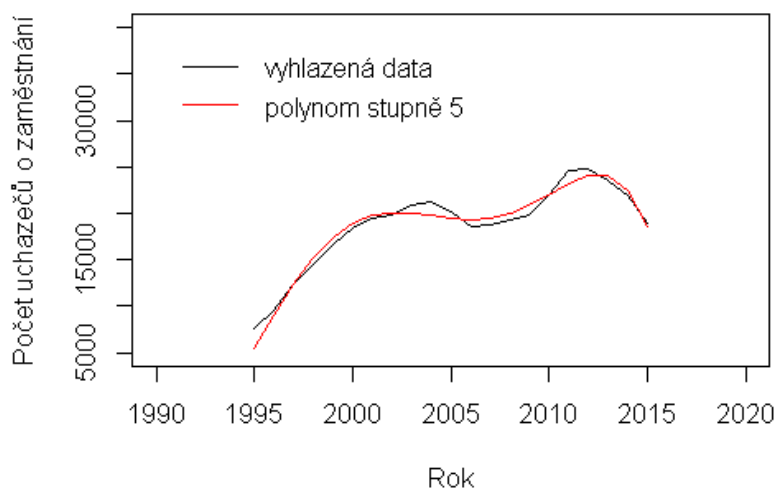
3.2.4 Proložení vyhlazených hodnot regresními polynomy

Znovu se budou hodnoty prokládat regresními polynomy. Je však nutné zaměřit se na hodnoty vyhlazené pomocí klouzavého průměru stupně dva. V tabulce je možné sledovat hodnoty koeficientu determinace u jednotlivých polynomů:

Tabulka 3.12: Analýza 2 - Výsledky proložení vyhlazené řady (klouzavými průměry) regresními polynomy.

Stupeň polynomu	1	2	3	4	5
Koeficient determinace	0.6654	0.8951	0.9306	0.9464	0.9806

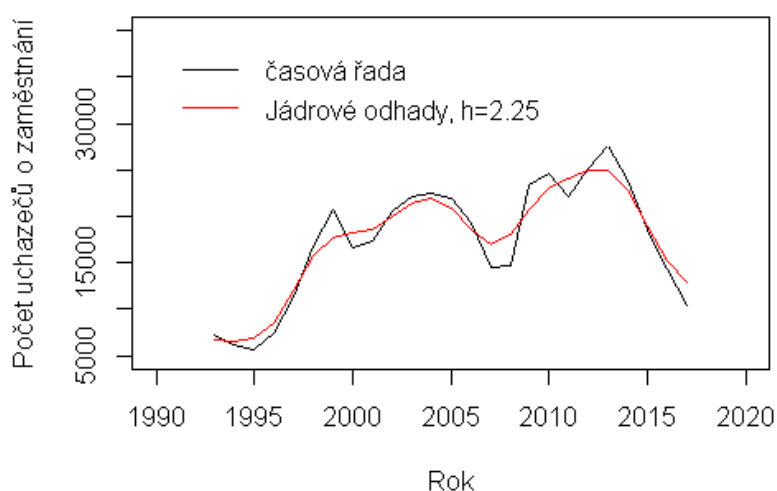
Už v tabulce 3.12 lze pozorovat, že vyhlazení dat regresním polynomem více odpovídá tvaru křivky. Toto tvrzení lze potvrdit pomocí grafického znázornění na obrázku 3.14, na kterém je možné vidět proložení polynomem stupně pět.



Obrázek 3.14: Graf vyhlazeného počtu uchazečů (klouzavým průměrem) proložený regresním polynomem stupně 5.

3.2.5 Využití jádrových odhadů

Důraz je kladen na druhý z neparametrických testů, kterým jsou jádrové odhady. Pro tuto časovou řadu byl zvolen jako nejvhodnější vyhlazovací parametr $h = 2.25$. Jak je vidět na obrázku 3.15, pomocí tohoto vyhlazovacího okna se zmírní růst zejména v roce 1999 a pokles v letech 2007 a 2008.



Obrázek 3.15: Graf počtu uchazečů proložený jádrovým odhadem s $h = 2.25$.

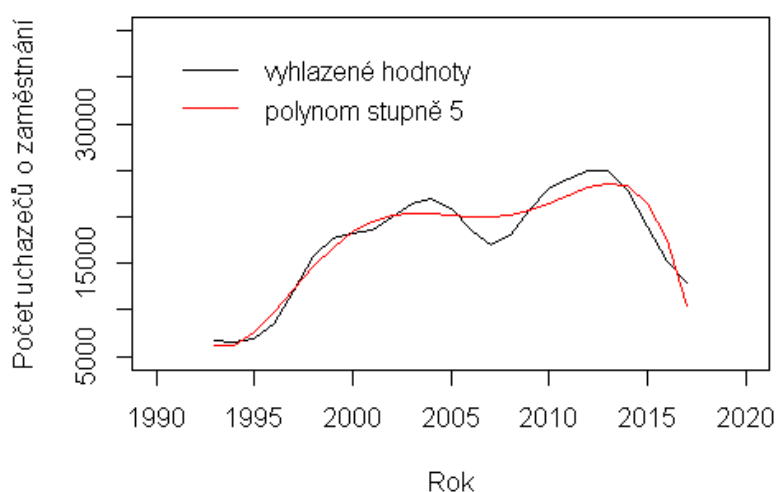
V následující tabulce je znázorněn koeficient determinace u upravených dat

s použitím regresního polynomu stupně jedna až pět.

Tabulka 3.13: Analýza 2 - Výsledky proložení vyhlazené řady (jádrovým odhadem) regresními polynomy

Stupeň polynomu	1	2	3	4	5
Koeficient determinace	0.4052	0.8022	0.8086	0.8458	0.9289

Z uvedených hodnot lze rozhodnout, že nejvhodnější proložení je pomocí polynomu stupně pět. Toto proložení je zobrazeno na obrázku 3.16.



Obrázek 3.16: Graf vyhlazeného počtu uchazečů (jádrovým odhadem) proložený regresním polynomelem stupně 5.

3.2.6 Shrnutí

Vypočtená data se znovu prozkoumala a porovnála se nejvhodnější proložení dat pomocí regresních polynomů. V tabulce 3.14 jsou vidět jednotlivé hodnoty koeficientu determinace u původních dat, u dat vyhlazených pomocí klouzavých průměrů a dat vyhlazených pomocí jádrových odhadů.

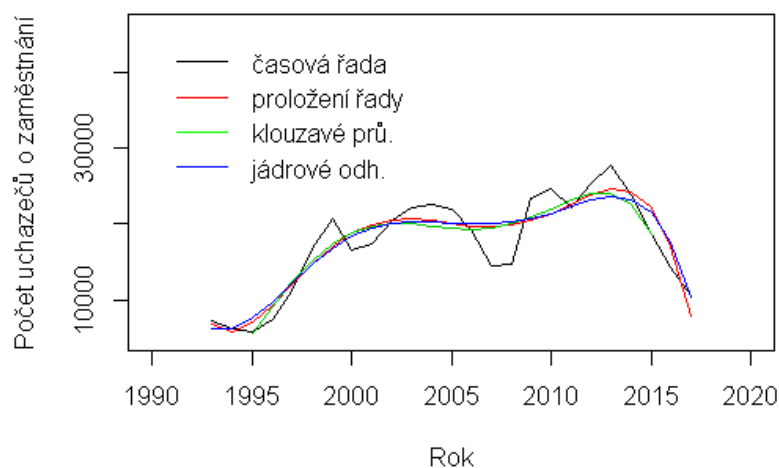
Pomocí celkového přehledu lze usoudit, že nejvhodnější proložení grafu je u vyhlazených hodnot pomocí klouzavých průměrů řádu dva (viz obrázek 3.14). U dat je hodnota koeficientu determinace ve výši 0.98057. Při využití jádrových odhadů s využitím vyhlazovacího okna ve velikosti 2.25 je proložení o 0.05168 horší. Není zcela jasné, zda by při použití jině širě vyhlazovacího okna, bylo proložení regresním polynomelem vhodnější než při použití klouzavých průměrů. Z této tabulky lze zjistit, že proložení polynomelem prvního stupně u tohoto příkladu je méně využitelné než u prvního příkladu. Důvod lze hledat v nestálosti tohoto

Tabulka 3.14: Analýza 2 - Tabulka všech vypočítaných hodnot

St. polynomu	Původní hodnoty	Vyhlazené klouzavými průměry	Vyhlazené jádrovými odhady
1	0.3082	0.6654	0.4052
2	0.6643	0.8951	0.8022
3	0.6764	0.9306	0.8086
4	0.7253	0.9464	0.8458
5	0.8421	0.9806	0.9289

grafu, u kterého se proložení pouze lineárním polynomem bude odrážet ve velké vzdálenosti jednotlivých bodů od přímky.

Na obrázku 3.17 je vidět proložení časové řady regresním polynomem, proložení regresním polynomem pokud se použijí klouzavé průměry a proložení polynomem s využitím jádrových odhadů.



Obrázek 3.17: Graf jednotlivých proložení časové řady s použitím různých metod.

3.3 Analýza dat o kurzu měny

Jako poslední se budou zkoumat data, se kterými se setkává minimálně jednou ročně každý, kdo jezdí na dovolenou nebo zahraniční cestu do zahraničí. Zkoumat se budou hodnoty vývoje kurzu české měny vůči euru za první čtvrtletí roku 2019.

Tato data vyhláší každý pracovní den Česká národní banka na svých stránkách. Z tohoto důvodu nelze nalézt žádný kurz ke dni, který vychází na víkend a ke dni, ve kterém je vyhlášen státní svátek. V tabulce 3.15 je možné pozorovat vývoj kurzu EUR/CZK v prvním čtvrtletí roku 2019. Hodnoty byly získány na internetovém portálu Kurzy.cz [6].

Tabulka 3.15: Kurz EUR/CZK za první čtvrtletí roku 2019.

Den	Kurz	Den	Kurz	Den	Kurz	Den	Kurz	Den	Kurz
2.1.	25.750	21.1.	25.595	7.2.	25.805	26.2.	25.665	15.3.	25.670
3.1.	25.680	22.1.	25.610	8.2.	25.805	27.2.	25.655	18.3.	25.615
4.1.	25.650	23.1.	25.695	11.2.	25.835	28.2.	25.600	19.3.	25.600
7.1.	25.575	24.1.	25.695	12.2.	25.870	1.3.	25.635	20.3.	25.645
8.1.	25.640	25.1.	25.695	13.2.	25.795	4.3.	25.625	21.3.	25.650
9.1.	25.630	28.1.	25.725	14.2.	25.790	5.3.	25.620	22.3.	25.725
10.1.	25.625	29.1.	25.745	15.2.	25.700	6.3.	25.595	25.3.	25.760
11.1.	25.605	30.1.	25.800	18.2.	25.715	7.3.	25.610	26.3.	25.770
14.1.	25.560	31.1.	25.760	19.2.	25.715	8.3.	25.640	27.3.	25.800
15.1.	25.570	1.2.	25.695	20.2.	25.680	11.3.	25.655	28.3.	25.780
16.1.	25.565	4.2.	25.730	21.2.	25.645	12.3.	25.670	29.3.	25.800
17.1.	25.535	5.2.	25.700	22.2.	25.665	13.3.	25.670		
18.1.	25.580	6.2.	25.785	25.2.	25.650	14.3.	25.670		

Tato časová řada se od prvních dvou liší v mnoha pohledech. Je zřejmé, že má mnohem více nasbíraných dat proti předchozím, ale přesto ji lze řadit do krátkodobých časových řad. Je to z důvodu, že jednotlivá pozorování jsou zjišťována v kratším časovém intervalu než je jeden rok. Dokonce ji lze zařadit do vysokofrekvenčních časových řad. Jedná se o finanční časovou řadu, která je získána z vývoje kurzu národní měny na finančních trzích.

3.3.1 Zkoumání elementárních charakteristik

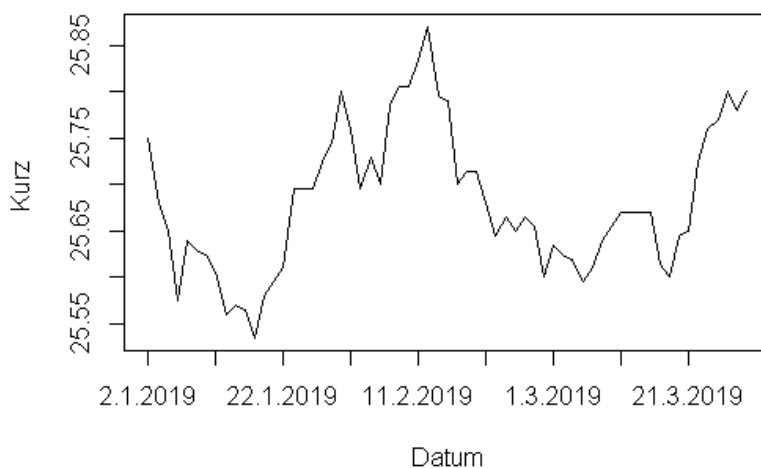
Z důvodu velkého počtu hodnot, se nevyepisují všechny vypočítané hodnoty elementárních charakteristik, nýbrž jsou zmíněny jen ty nejzajímavější.

První diference říká, že prvních deset pozorování má spíše klesající tendenci. Do 29.1.2019 pozorování naopak stoupají, pak přichází mírný pokles a data znovu stoupají až do 12.2.2019. Poté přichází velké snížení během tří dnů, ve

kterém by se zaplatilo o 0,17 Kč za euro méně. Do 45 hodnoty pozprování kurz střídavě klesá a roste. Ke konci časové řady je možné říci, že kurz spíše roste, ojedinele se objevil pokles dne 20. 3. 2019.

Hodnota koeficientu růstu se pohybuje od 0.99651 do 1.00332 a znamená, o kolik se zvýšil či snížil kurz měny. Průměrný koeficient růstu je roven 1.

Na obrázku 3.18 je znázorněna celková struktura řady kurzů.



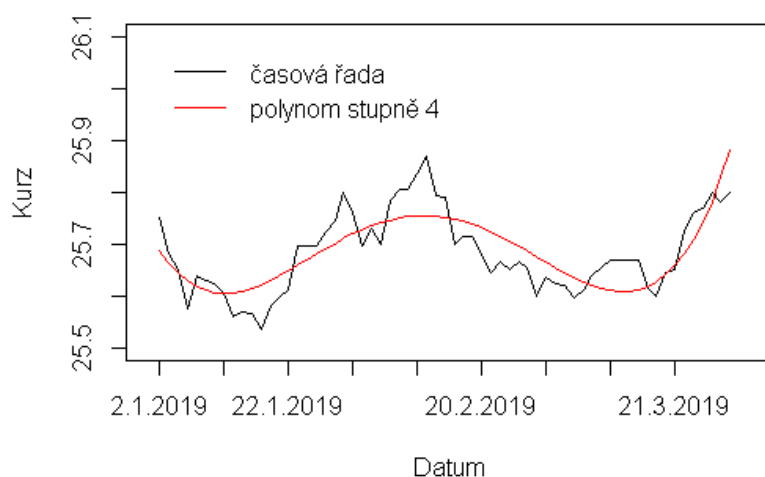
Obrázek 3.18: Graf kurzu EUR/CZK za první čtvrdletí roku 2019.

3.3.2 Proložení časové řady regresním polynomem

V tabulce 3.16 jsou vidět hodnoty koeficientu determinace u polynomů stupně jedna až čtyři a na obrázku 3.19 je vidět proložení polynomem stupně čtyři, který byl zvolen za nejvhodnější.

Tabulka 3.16: Analýza 3 - Výsledky proložení řady regresními polynomy.

Stupeň polynomu	1	2	3	4
Koeficient determinace	0.0439	0.0679	0.2103	0.6314



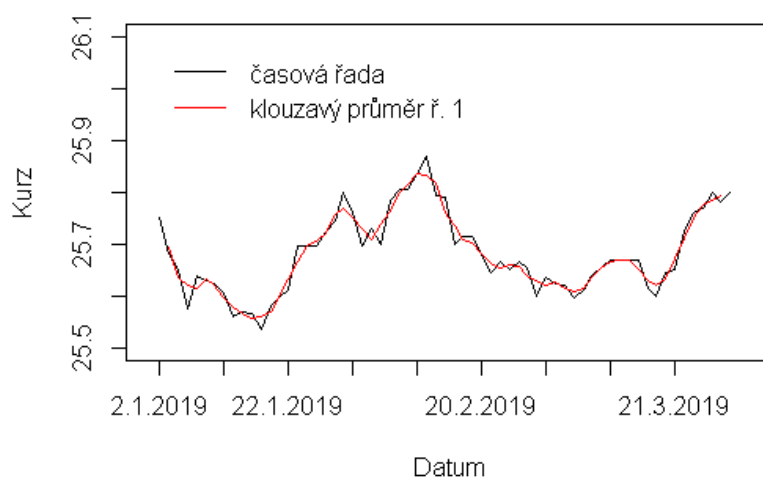
Obrázek 3.19: Graf kurzu proložený regresním polynomem stupně 4.

3.3.3 Vyhlazení časové řady pomocí klouzavých průměrů

Znovu byly vypočítány hodnoty V_k pro určení vhodného stupně klouzavého průměru a jejich výsledky jsou vidět v tabulce 3.17. Od druhého stupně hodnoty začínají konvergovat a proto je zvolen klouzavý průměr stupně jedna viz obrázek 3.20.

Tabulka 3.17: Analýza 3 - Vypočítané hodnoty V_k pro jednotlivé stupně vyhlazení.

k	1	2	3	4	5	6
V_k	7.1653	4.8060	4.5917	4.5917	4.4211	4.2414



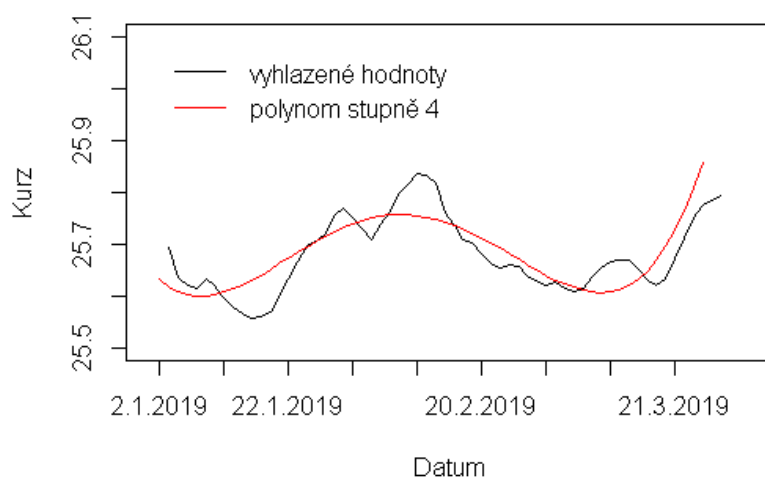
Obrázek 3.20: Graf kurzu vyhlazený pomocí klouzavého průměru stupně 1.

3.3.4 Proložení vyhlazených hodnot regresním polynommem

Po vyhlazení kurzu byly vypočítané hodnoty znovu prokládány regresními polynomy. V tabulce 3.18 jsou zobrazeny koeficienty determinace u různých stupňů polynomu. Pokud budou hodnoty porovnány s původními, je možné sledovat zlepšení proložení u všech stupňů. Grafické znázornění pozorujeme na obrázku 3.21.

Tabulka 3.18: Analýza 3 - Výsledky proložení vyhlazené řady (klouzavými průměry) regresními polynomy.

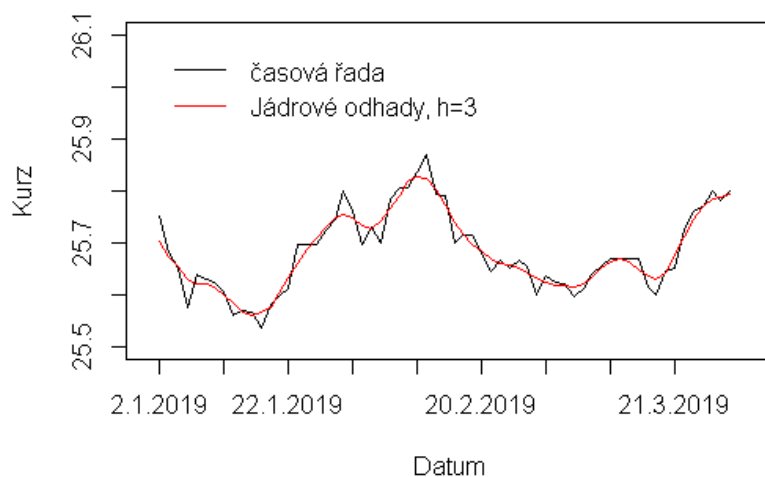
Stupeň polynomu	1	2	3	4
Koeficient determinace	0.0445	0.1159	0.3171	0.7105



Obrázek 3.21: Graf vyhlazeného kurzu (klouzavým průměrem) proložený regresním polynomem stupně 4.

3.3.5 Využití jádrových odhadů

Pro vyhlazení pomocí jádrových odhadů bylo využito vyhlazovací okno velikosti tři. Tato operace je znázorněna na obrázku 3.22.



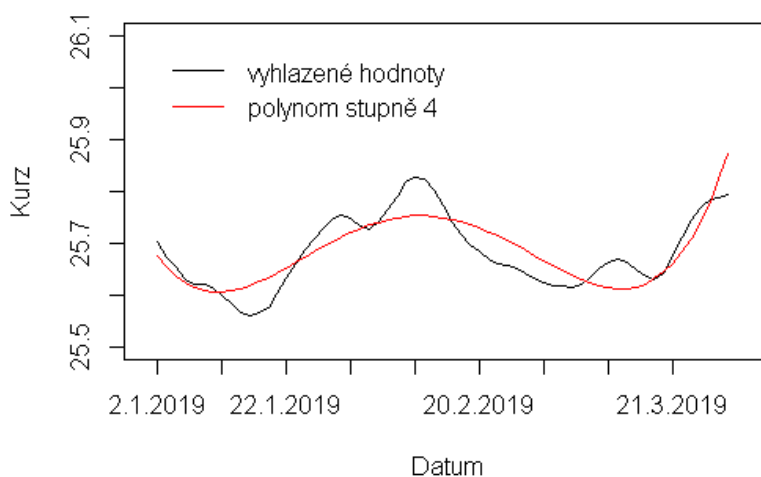
Obrázek 3.22: Graf kurzu proložený jádrovým odhadem s $h = 3$.

Nyní se nové hodnoty proloží regresními polynomy a zjistí se jejich hodnoty koeficientu determinace. Stupeň polynomu, který má nejbližší koeficient determinace k jedné a jeho graf nejvíce odpovídá časové řadě, bude zvolen jako

nejvhodnější. V tabulce 3.19 jsou zapsány hodnoty koeficientu.

Tabulka 3.19: Analýza 3 - Výsledky proložení vyhlazené řady (jádrovým odhadem) regresními polynomy.

Stupeň polynomu	1	2	3	4
Koeficient determinace	0.0559	0.0906	0.2689	0.7156



Obrázek 3.23: Graf vyhlazeného kurzu (jádrovým odhadem) proložený regresním polynomem stupně 4.

3.3.6 Shrnutí

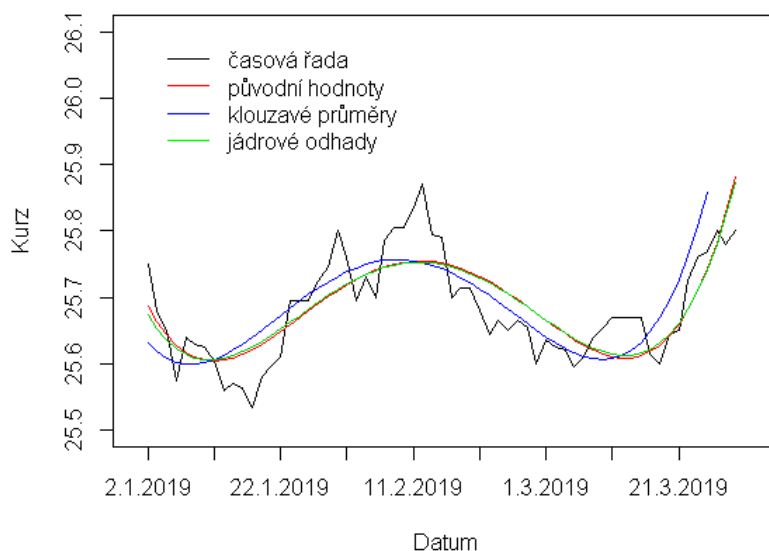
Nakonec se prozkoumají všechny hodnoty koeficientu determinace najednou. V tabulce 3.20 jsou vyznačeny všechny hodnoty koeficientu u jednotlivých stupňů polynomů, při využití všech uvedených metod.

Tabulka 3.20: Analýza 3 - Tabulka všech vypočítaných hodnot

St. polynomu	Původní hodnoty	Vyhlazené klouzavými průměry	Vyhlazené jádrovými odhady
1	0.0439	0.0445	0.0559
2	0.0679	0.1159	0.0906
3	0.2103	0.3171	0.2689
4	0.6314	0.7105	0.7156

Z přehledu hodnot lze usoudit, že nejvhodnějším zvoleným proložením bude polynom stupně čtyři u hodnot vyhlazených pomocí jádrových odhadů. Hodnota

koeficientu determinace je ve výši 0.71559. Tato hodnota je jen o 0.005 lepší nežli proložení pomocí regresního polynomu s použitím klouzavých průměrů. V grafu 3.24 jsou vykresleny polynomy stupně čtyři, při zadaných hodnotách časové řady, při využití klouzavých průměrů a jádrových odhadů.



Obrázek 3.24: Graf všech druhů proložení časové řady.

3.4 Tabulka vypočítaných hodnot

V tabulce 3.21 jsou zaneseny jednotlivé hodnoty koeficientu determinace vypočítané pro všechna proložení. Sloupec označený písmenem p udává proložení původních hodnot polynomem různého stupně, písmeno k označuje proložení vyhlazených hodnot pomocí klouzavých průměrů polynomem různého stupně a j proložení vyhlazených hodnot pomocí jádrových odhadů polynomem různého stupně.

Tabulka 3.21: Celkové porovnání hodnot koeficientu determinace

Stupeň polynomu	Analýza 1			Analýza 2			Analýza 3		
	p	k	j	p	k	j	p	k	j
1	0.957	0.958	0.962	0.308	0.665	0.405	0.043	0.044	0.056
2	0.979	0.985	0.982	0.664	0.895	0.802	0.067	0.115	0.090
3	0.991	0.992	0.995	0.676	0.930	0.808	0.210	0.317	0.268
4	0.995	0.997	0.998	0.725	0.946	0.845	0.631	0.710	0.715
5	0.996	0.997	0.999	0.842	0.980	0.928			

Závěr

Hlavním cílem této bakalářské práce bylo zanalyzování vybraných časových řad. Příklady, které jsou v této práci uvedeny, mají popisovat různé druhy ekonomických časových řad. V závěru jsou uvedeny jednotlivé výsledky pozorování a provedeno jejich shrnutí.

Mezi prvními rozdíly u časových řad, o kterých se zde píše, jsou různé doby pozorování řady. První a druhý příklad jsou řazeny do dlouhodobých časových řad (data byla pozorována po několik let) a poslední příklad je uveden jako krátkodobá časová řada či dokonce vysokofrekvenční časová řada (data byla sledována po dobu 3 měsíců). Jako další rozdíl mezi příklady lze uvést charakter jednotlivých dat. U prvního příkladu se jedná o intervalové časové řady, u dalších dvou o řady okamžikové. První časová řada má rostoucí charakter a druhé dvě jsou proměnlivé.

Z tabulky 3.21 je vidět, že v prvním příkladu se hodnota koeficientu pohybovala od 0.957 do 0.999, u druhého příkladu od 0.308 do 0.980 a ve třetím od 0.043 do 0.715. Z těchto výsledků lze usoudit, že čím je časová řada pravidelnější, tím více se koeficient determinace blíží k jedné.

Dále je patrné, že vyhlazené hodnoty více odpovídají proložením všech druhů stupňů polynomů. U prvního příkladu je vidět, že vhodnost proložení se střídá mezi klouzavými průměry a jádrovými odhady. V druhém příkladu je zřejmé, že vhodnější ve všech stupních polynomů je využití klouzavých průměrů. V posledním příkladu je koeficient až do třetího stupně polynomu blíže k jedné při využití klouzavých průměrů. U stupně číslo čtyři je lepší použití jádrových odhadů.

Předmětem dalšího zkoumání by mohlo být určení vhodného vyhlazovacího parametru. Pokud by byl tento parametr nalezen pro uvedené příklady, nemuselo by být využití klouzavých průměrů vhodnější než jádrové odhady.

Seznam tabulek

2.1	Ukázky různých jader.	24
3.1	Počet domácností s vlastním PC.	26
3.2	Analýza 1 - Základní elementární charakteristiky.	27
3.3	Analýza 1 - Výsledky proložení řady regresními polynomy.	28
3.4	Analýza 1 - Vypočítané hodnoty V_k pro jednotlivé stupně vyhlazení.	30
3.5	Analýza 1 - Výsledky proložení vyhlazené řady (klouzavými průměry) regresními polynomy.	31
3.6	Analýza 1 - Výsledky proložení vyhlazené řady (jádrovým odhadem) regresními polynomy.	34
3.7	Analýza 1 - Tabulka všech vypočítaných hodnot.	35
3.8	Počet uchazečů o zaměstnání v Královéhradeckém kraji.	36
3.9	Analýza 2 - Základní elementární charakteristiky.	37
3.10	Analýza 2 - Výsledky proložení řady regresními polynomy.	38
3.11	Analýza 2 - Vypočítané hodnoty V_k pro jednotlivé stupně vyhlazení.	38
3.12	Analýza 2 - Výsledky proložení vyhlazené řady (klouzavými průměry) regresními polynomy.	39
3.13	Analýza 2 - Výsledky proložení vyhlazené řady (jádrovým odhadem) regresními polynomy	41
3.14	Analýza 2 - Tabulka všech vypočítaných hodnot	42
3.15	Kurz EUR/CZK za první čtvrtletí roku 2019.	43
3.16	Analýza 3 - Výsledky proložení řady regresními polynomy.	44
3.17	Analýza 3 - Vypočítané hodnoty V_k pro jednotlivé stupně vyhlazení.	45
3.18	Analýza 3 - Výsledky proložení vyhlazené řady (klouzavými průměry) regresními polynomy.	46
3.19	Analýza 3 - Výsledky proložení vyhlazené řady (jádrovým odhadem) regresními polynomy.	48
3.20	Analýza 3 - Tabulka všech vypočítaných hodnot	48
3.21	Celkové porovnání hodnot koeficientu determinace	49

Seznam obrázků

1.1	Grafické znázornění vzorové časové řady.	11
1.2	Dekompozice časové řady $Y(t)$. Červené křivky naznačují tvary trendové, sezonní, cyklické a náhodné složky.	12
1.3	Tvar trendu u vzorové časové řady.	13
1.4	Tvar sezonní složky u vzorové časové řady.	15
1.5	Tvar cyklické složky u vzorové časové řady.	16
1.6	Tvar náhodné složky u vzorové časové řady.	16
2.1	Ukázka klouzavých průměrů.	22
3.1	Graf počtu domácností s vlastním osobním počítačem.	28
3.2	Graf domácností proložený regresním polynomem stupně 1.	29
3.3	Graf domácností proložený regresním polynomem stupně 5.	29
3.4	Graf domácností proložený klouzavým průměrem délky 8.	30
3.5	Graf domácností proložený klouzavým průměrem stupně 1.	31
3.6	Graf vyhlazeného počtu domácností (klouzavým průměrem) proložený regresním polynomem stupně 4.	32
3.7	Graf domácností proložený jádrovým odhadem s $h = 1.5$	33
3.8	Graf domácností proložený jádrovým odhadem s $h = 4$	33
3.9	Graf vyhlazeného počtu domácností (jádrovým odhadem) proložený regresním polynomem stupně 5.	34
3.10	Graf proložení časové řady s použitím různých metod.	35
3.11	Graf počtu uchazečů o zaměstnání.	37
3.12	Graf počtu uchazečů proložený regresním polynomem stupně 5.	38
3.13	Graf počtu uchazečů vyhlazený pomocí klouzavého průměru stupně 2.	39
3.14	Graf vyhlazeného počtu uchazečů (klouzavým průměrem) proložený regresním polynomem stupně 5.	40
3.15	Graf počtu uchazečů proložený jádrovým odhadem s $h = 2.25$	40
3.16	Graf vyhlazeného počtu uchazečů (jádrovým odhadem) proložený regresním polynomem stupně 5.	41
3.17	Graf jednotlivých proložení časové řady s použitím různých metod.	42
3.18	Graf kurzu EUR/CZK za první čtvrtletí roku 2019.	44
3.19	Graf kurzu proložený regresním polynomem stupně 4.	45

3.20	Graf kurzu vyhlazený pomocí klouzavého průměru stupně 1.	46
3.21	Graf vyhlazeného kurzu (klouzavým průměrem) proložený regresním polynomem stupně 4.	47
3.22	Graf kurzu proložený jádrovým odhadem s $h = 3$	47
3.23	Graf vyhlazeného kurzu (jádrovým odhadem) proložený regresním polynomem stupně 4.	48
3.24	Graf všech druhů proložení časové řady.	49

Seznam použité literatury

- [1] ANDĚL, Jiří. *Statistické metody*. 4., upr. vyd. Praha: Matfyzpress, 2007, 299 s. ISBN 978-80-7378-003-8.
- [2] ARLT, Josef a Markéta ARLTOVÁ. *Ekonomické časové řady*. Praha: Grada, 2007, 285 s. ISBN 80-247-1319-5.
- [3] ARLT, Josef a Markéta ARLTOVÁ. *Finanční časové řady*. Praha: Grada, 2003, 220 s. Expert. ISBN 80-247-0330-0.
- [4] CIPRA, Tomáš. *Analýza časových řad s aplikacemi v ekonomii*. Praha: Státní nakladatelství technické literatury, 1986, 246 s.
- [5] CIPRA, Tomáš. *Finanční ekonometrie*. Praha: Ekopress, 2008, 538 s. ISBN 978-80-86929-43-9.
- [6] Česká národní banka, Kurzycz [online]. [Cit.30.3.2019]. Dostupné z: <https://www.kurzy.cz/kurzy-men/grafy/CZK-EUR/>
- [7] Český statistický úřad [online]. [Cit.11.3.2019]. Dostupné z: <https://www.czso.cz/csu/czso/ceska-republika-od-roku-1989-v-cislech-2017-24bfnixod8>
- [8] HOLČÍK, Jiří, KOMENDA, Martin (eds.) a kol. *Matematická biologie: e-learningová učebnice*. [online]. [Cit.12.2.2019]. Dostupné z: <http://portal.matematickabiologie.cz/index.php?pg=analiza-a-hodnoceni-biologicky-ch-dat-teorie-a-praxe-jadroveho-vyhlazovani>
- [9] KOZÁK, Josef, Josef ARLT a Richard HINDLS. *Úvod do analýzy ekonomických časových řad*. Praha: VŠE, 1994, 207 s. ISBN 80-7079-760-6.
- [10] SEGER, Jan a Richard HINDLS. *Statistické metody v tržním hospodářství*. Praha: Victoria Publishing, 1995, 435 s. ISBN 80-7187-058-7.