

Jihočeská univerzita v Českých Budějovicích
Ekonomická fakulta
Katedra aplikované matematiky a informatiky

Bakalářská práce

Analýza finančních dat metodami prostorů funkcí

Vypracoval: Adam Huleš

Vedoucí práce: Mgr. Petr Chládek, Ph.D.

České Budějovice, 2024

JIHOČESKÁ UNIVERZITA V ČESKÝCH BUDĚJOVICÍCH

Ekonomická fakulta
Akademický rok: 2023/2024

ZADÁNÍ BAKALÁŘSKÉ PRÁCE

(projektu, uměleckého díla, uměleckého výkonu)

Jméno a příjmení: Adam HULEŠ
Osobní číslo: E21134
Studijní program: B0688A140010 Podniková informatika
Téma práce: Analýza finančních dat metodami prostorů funkcí
Zadávající katedra: Katedra aplikované matematiky a informatiky

Zásady pro vypracování

Matematická analýza vyvinula celou řadu způsobů, jak studovat charakter funkce. Nejtypičtější se měří regularita, tedy hladkost, a integrabilita, což je zobecnění omezenosti. Student si nastuduje základní pojmy prostorů funkcí a následně bude pro vybraná data z finančního sektoru zkoumat povahu dat z tohoto hlediska. Na základě těchto zkoumání bude uvažováno, jakým vhodným způsobem neúplná data doplnit mezi měřené hodnoty, interpolovat, nebo, po sekvenci měřených hodnot, extrapolovat.

Metodický postup:

1. Student se seznámí s teorií Sobolevových prostorů v jedné dimenzi a technikami interpolace a extrapolace.
2. Student navrhne experiment pro ověření kvality interpolační metody jako odhadu funkce.
3. Student srovná výsledky interpolačních metod s danou hustotou uzlů pro různá data a změří chyby jednotlivých nastavení.
4. Student vysloví doporučení, pro jaké typy dat jsou které metody vhodné a které metody jsou zcela nevhodné.

Rozsah pracovní zprávy: 30 – 50 str.
Rozsah grafických prací: dle potřeby
Forma zpracování bakalářské práce: tištěná

Seznam doporučené literatury:

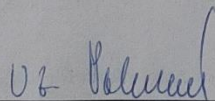
1. Hošek, R. (2016). *Společný jazyk matematiky a ekonomie*. (Bakalářská práce). Jihočeská univerzita v Českých Budějovicích, Ekonomická fakulta.
2. Chow, G. C., & Lin, A. L. (1971). Best linear unbiased interpolation, distribution, and extrapolation of time series by related series. *The review of Economics and Statistics*, 372-375.
3. Leoni, G. (2017). *A first course in Sobolev spaces*. American Mathematical Soc.
4. Pavia-Miralles, J. M. (2010). A survey of methods to interpolate, distribute and extra-polate time series. *Journal of Service Science and Management*, 3(04), 449.

Vedoucí bakalářské práce: Mgr. Petr Chládek, Ph.D.
Katedra aplikované matematiky a informatiky

Konzultant bakalářské práce: **RNDr. Tomáš Roskovec, Ph.D.**
Katedra aplikované matematiky a informatiky

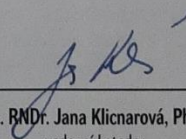
Datum zadání bakalářské práce: **19. prosince 2023**

Termín odevzdání bakalářské práce: **14. dubna 2024**



doc. RNDr. Zuzana Dvořáková Lišková, Ph.D.
děkanka

JIHOČESKÁ UNIVERZITA
V ČESKÝCH BUDĚJOVICÍCH
EKONOMICKÁ FAKULTA
Studentská 13 (23)
370 05 České Budějovice



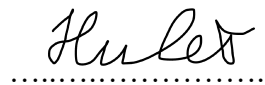
doc. RNDr. Jana Klicnarová, Ph.D.
vedoucí katedry

V Českých Budějovicích dne 19. prosince 2023

Prohlašuji, že svou bakalářskou práci jsem vypracoval samostatně pouze s použitím pramenů a literatury uvedených v seznamu citované literatury.

Prohlašuji, že v souladu s § 47b zákona č. 111/1998 Sb. v platném znění souhlasím se zveřejněním své bakalářské práce, a to v nezkrácené podobě elektronickou cestou ve veřejně přístupné části databáze STAG provozované Jihočeskou univerzitou v Českých Budějovicích na jejich internetových stránkách, a to se zachováním mého autorského práva k odevzdanému textu této kvalifikační práce. Souhlasím dále s tím, aby toutéž elektronickou cestou byly v souladu s uvedeným ustanovením zákona č. 111/1998 Sb. zveřejněny posudky školitele a oponentů práce i záznam o průběhu a výsledku obhajoby kvalifikační práce. Rovněž souhlasím s porovnáním textu mé kvalifikační práce s databází kvalifikačních prací Theses.cz provozovanou Národním registrem vysokoškolských kvalifikačních prací a systémem na odhalování plagiátů.

V Českých Budějovicích dne 11. dubna 2024



Adam Huleš

Na tomto místě bych rád poděkoval panu RNDr. Tomáši Roskovcovi, Ph.D. za cenné rady, vstřícnost a trpělivost, jež mi věnoval po celou dobu tvorby této práce. Velký dík patří též mé sestře, která mi byla vzorem na mé akademické cestě.

Obsah

1	Úvod	8
1.1	Cíle práce	8
2	Spojitosť a hladkost funkce.....	10
2.1	Spojitosť funkce.....	10
2.2	Hladkost funkce	10
2.2.1	Příklady hladkých funkcí	10
2.2.2	Příklady nehladkých či nespojitých funkcí.....	10
2.3	Weierstrassova věta.....	11
3	Základní typy (kvantitativních) dat	12
3.1	Diskrétní data	12
3.2	Spojité data.....	12
4	Interpolace dat	14
4.1	Po částech konstantní funkce	14
4.2	Po částech lineární funkce.....	14
4.3	Po částech polynomiální funkce.....	14
4.3.1	Kubický interpolační spline	14
4.3.2	Lagrangeova metoda – konstrukce interpolačního polynomu.....	18
5	Fourierovy řady	20
6	Regrese	22
6.1	Požadavky na regresní křivku	23
7	Extrapolace	24
7.1	Využití extrapolace při konstrukci kubického splinu	24
8	Měření chyby.....	26
8.1	Metoda nejmenších čtverců.....	26
8.2	Lebesgueovy prostory	28
8.3	Sobolevovy prostory	29

9	Aplikace interpolačních a regresních metod na měnové kurzy	31
9.1	Interpolace.....	31
9.2	Lineární regrese.....	34
9.3	Polynomiální regrese.....	35
9.4	Regrese Fourierovou řadou	38
9.5	Regrese v čase neobvyklých událostí.....	39
10	Aplikace interpolačních a regresních metod na vývoj počtu obyvatel	42
10.1	Interpolace.....	42
10.2	Lineární regrese.....	43
10.3	Polynomiální regrese.....	44
10.4	Regrese Fourierovou řadou	45
10.5	Regrese v čase neobvyklých událostí.....	46
	Závěr	48
	Summary.....	49
	Seznam použité literatury	50
	Seznam obrázků a tabulek	51
	Přílohy.....	52
A	Měsíční měnový kurz devizového trhu EUR/CZK.....	52
B	Čtvrtletní měnový kurz devizového trhu EUR/CZK	53
C	Roční měnový kurz devizového trhu EUR/CZK.....	54
D	Předpisy kubického interpolačního splinu	55
E	Roční měnový kurz devizového trhu EUR/RUB.....	57
F	Kód pro zjištění vhodného počtu harmonických složek Fourierovy řady	57
G	Kód pro regresi Fourierovou řadou	58
H	Vývoj počtu obyvatel ČR	59

1 Úvod

Je lidskou přirozeností snažit se na základě vlastní zkušenosti odhadnout další události. Nejinak je tomu v ekonomii, kde takový správný odhad může znamenat rozdíl mezi milionovými zisky a pro podnik likvidační ztrátou prostředků. Stalo se proto dobrou praxí, že podniky využívají analytických metod ke zkoumání dosavadního chování trhu a jeho různých aspektů v daném odvětví, ať už se jedná o vývoj měnových kurzů, cen akcií nebo prodejnost výrobku v různých věkových skupinách.

Z technologického hlediska není dost dobře možné docílit nepřetržitého pozorování a zaznamenávání dění na trhu. Navíc každé takové pozorování stojí nemalé náklady, které s intenzitou pozorování velice rychle rostou. Přitom je přirozené, že čím více informací o zkoumaném problému máme, tím lepší máme předpoklady se vhodně rozhodnout. Jaké množství dat je tedy pro jaké účely dostatečné? Jaké nejmenší množství informací nám stačí k tomu, abychom dokázali s určitou pravděpodobností předvídat, jak se datová sada chovala v době, kdy měření neprobíhala? Jedním ze způsobů odhadu budoucího chování naměřených dat je kombinace matematicko-analytických metod interpolace, extrapolace a regrese.

Interpolace nám nabízí doplnění naměřených dat v celém časovém intervalu zkoumání, zatímco extrapolace umožňuje rozšířit získané informace mimo tento interval (do minulosti i do budoucnosti). Regrese nám pak poskytuje způsoby, jak získané informace zjednodušit, aby bylo pro člověka snazší s nimi pracovat, a především jim porozumět. Regresní metody se také dají využít při odhadu chování sledované veličiny, a to jak uvnitř měřeného intervalu, tak mimo něj (opět v minulosti i budoucnosti).

1.1 Cíle práce

1. Seznámit čtenáře s technikami interpolace, extrapolace a regrese a způsoby měření chyb jednotlivých metod.
2. Navrhnout experiment pro ověření kvality interpolační a regresní metody jako odhadu chování funkce popisující vývoj vybraných ekonomických dat.
3. Srovnat výsledky interpolačních metod s danou hustotou uzlů pro různá data a změřit chyby jednotlivých nastavení.
4. Stanovit doporučení, pro jaké typy dat jsou které metody vhodné a které metody jsou zcela nevhodné.

Autor si je vědom existence množství kvalitních knih a publikací, které podrobně popisují analýzu dat matematickými metodami. Tyto jsou však často svým rozsahem a odborností nad rámec znalostí studentů ekonomie a věnují se problematice převážně z matematického hlediska.

Cílem této práce je využití znalostí těchto metod k vytvoření přesahu a propojení s oborem ekonomie skrze jejich využití na reálných ekonomických datech. Na datové sady různých měnových kurzů aplikujeme několik regresních funkcí, porovnáme jejich přesnost z hlediska odhadu budoucího chování dat a pokusíme se nalézt hlavní faktory, které umožňují (či znemožňují) jejich provedení a ovlivňují přesnost jejich odhadu.

2 Spojitost a hladkost funkce

2.1 Spojitost funkce

Necht' f je funkce definovaná na nějakém okolí bodu a . Řekneme, že funkce f je spojitá v a , jestliže f konverguje v a k $f(a)$, nebo

$$\lim_{x \rightarrow a} f(x) = f(a). \quad (1)$$

Řekneme, že funkce f je spojitá v otevřeném intervalu (a, b) , je-li spojitá v každém bodě tohoto intervalu.

2.2 Hladkost funkce

Bud' M otevřená množina. Řekneme, že funkce f je hladká na M , jestliže má na množině M spojitě všechny derivace prvního řádu

$$\lim_{x \rightarrow x_0} f'(x) = f'(x_0), \forall x_0 \in M. \quad (2)$$

Z výše uvedeného vyplývá, že hladkost je nadřazenou vlastností spojitosti. Tj. všechny hladké funkce jsou vždy spojitě, zatímco ne všechny spojitě funkce jsou hladké.

2.2.1 Příklady hladkých funkcí

Mezi nejznámější a nejpoužívanější hladké funkce patří:

- polynomy všech řádů,
- exponenciální funkce,
- funkce sinus a kosinus,
- funkce $\frac{1}{x^2+1}$.

2.2.2 Příklady nehladkých či nespojitých funkcí

- Absolutní hodnota – nemá derivaci v nulových bodech, tedy není hladká,
- po částech lineární funkce – nemá derivaci v bodech napojení lineárních částí,
- statistický šum,
- Brownův pohyb.

2.3 Weierstrassova věta

Při analýze reálných dat je velmi často užitečné nalézt lokální extrém (tj. maximum a minimum) zkoumané funkce. Jelikož v ekonomii pracujeme nejčastěji s daty v jedné dimenzi, můžeme zde uplatnit Weierstrassovu větu, která říká, že pokud je funkce f spojitá na uzavřeném intervalu $[a, b]$, nabývá na tomto intervalu své minimální i maximální hodnoty (Šťastný, 2011).

Tyto extrémní hodnoty nalezneme mezi stacionárními body funkce a krajními body intervalu $[a, b]$. Říkáme, že bod x_i je stacionárním bodem funkce $f(x)$, jestliže platí:

$$f'(x_i) = 0. \quad (3)$$

Zjistíme-li hodnoty druhých derivací funkce $f(x)$ ve všech těchto bodech, pak platí, že pokud $f''(x_i) > 0$, x_i je minimem funkce $f(x)$ a pokud $f''(x_i) < 0$, x_i je maximem $f(x)$.

3 Základní typy (kvantitativních) dat

Data jsou tvarem množného čísla latinského slova *datum*, které lze vyjádřit jako *něco daného* a které bylo původně odvozeno z přičestí minulého slova *dare*, teda *dát*. Lze je tedy chápat jako jednoznačnou reprezentaci faktů, atributů, odrazu dějů a věcí (Sklenák, 2001).

Ve statistice jsou data definována jako fakta a údaje shromážděné společně podle věcné souvislosti pro účely zkoumání a analýzy.

3.1 Diskrétní data

Diskrétní data jsou taková data, která mohou nabývat pouze spočetně mnoha hodnot. Při číselné reprezentaci jsou taková data na reálné ose zobrazena pomocí izolovaných bodů. Můžeme si je také představit jako spočetné množiny prvků. Praktickým příkladem diskretních dat tak může být sčítání populace, které probíhá každý rok.

Mezi nejvýznamnější výhody diskretního pojetí dat patří:

- Snadné získávání dat – jsme schopni zjistit/změřit konkrétní hodnotu ve vybraném časovém okamžiku, případně měření opakovat v pravidelných časových intervalech.
- Snadné ukládání – samostatné hodnoty lze snadno uchovat ve formě tabulky či seznamu. Toto pojetí navíc umožňuje počítačům uchovávat přesně všechny získané hodnoty.

Největší nevýhodou diskretních dat je:

- Nepřesnost měření – měřením jsme sice schopni získat přesné hodnoty v daném časovém okamžiku, ale tyto často nerepresentují realitu se 100% přesností. V důsledku lidských a přístrojových nedostatků tak může docházet k významnému zkreslení hodnot, které pak ovlivňuje výsledky naší analýzy.

3.2 Spojitá data

Spojitá data jsou taková data, která mohou nabývat jakýchkoliv hodnot v určitém rozmezí (intervalu). To znamená, že mezi jednotlivými hodnotami existuje nekonečně mnoho možných hodnot, což je typické pro data, která lze reprezentovat na spojitém číselném

měřítku. Příklady spojitých dat zahrnují veličiny jako venkovní teplota, lidský věk a podobně.

Mezi nejvýznamnější výhody spojitého pojetí dat patří především:

- Snadná interpretace – spojitě funkce umožňují zjištění funkční hodnoty v jakémkoliv bodě popisovaného intervalu a snadné použití analytických metod.

Největšími nevýhodami spojitých dat jsou:

- Obtížné získávání – v praxi je v podstatě nemožné získávat reálné hodnoty nepřetržitě. Např. teplotu nelze zaznamenávat v reálném čase jinak než diskrétně.
- Obtížné ukládání – máme-li data popsána spojitou funkcí, uchovává se v počítačích tato funkce popsána pouze pomocí parametrů – tedy diskrétních dat.

Statistické a numerické metody téměř výhradně pracují s diskrétními údaji. Pokud je potřeba pracovat se spojitými, z diskrétních dat je modelují. Přestože spojitá data by lépe umožnila interpretovat i predikovat zkoumaný jev, sebraná a uložená data jsou diskrétní. Se spojitými daty pracujeme v teoretických úvahách a modelech, ale je prakticky nemožné je změřit.

4 Interpolace dat

Interpolací rozumíme nalezení přibližného tvaru funkce v určitém intervalu na základě několika jejích známých (naměřených) hodnot. Jedná se tedy aproximaci diskrétní funkce funkcí spojitou, což umožňuje následně provádět podrobnější analýzu, než by na samotných empirických datech bylo možné. Existuje několik způsobů, jak stanovit funkční hodnoty $f(x_0), f(x_1), \dots, f(x_n)$ všech bodů x_0, x_1, \dots, x_n zkoumaného intervalu.

4.1 Po částech konstantní funkce

Nejjednodušším způsobem, jak vyplnit všechny hodnoty v určitém intervalu z několika známých bodů je po částech konstantní funkce. Pro tu platí:

$$f(x) = f(x_i) \text{ pro } x_i \leq x < x_{i+1}, \quad (4)$$

kde $f(x_i)$ jsou známé hodnoty funkce v uzlech měření x_i . Taková funkce však není spojitá, proto není jako reprezentace dat z hlediska dalších operací příliš vhodná.

4.2 Po částech lineární funkce

Chceme-li zajistit spojitost funkce, a přitom zachovat jednoduchost výpočtu funkčních hodnot, lze počáteční data proložit po částech lineární (lomenou) funkcí. Pro tu platí:

$$f(x) = \frac{f(x_{i+1}) - f(x_i)}{x_{i+1} - x_i} \cdot (x - x_i) + f(x_i) \text{ pro } x_i \leq x \leq x_{i+1}. \quad (5)$$

Tato funkce je již spojitá, ale v bodech napojení není hladká, což stále znemožňuje některé analytické operace.

4.3 Po částech polynomiální funkce

4.3.1 Kubický interpolační spline

Spline je po částech polynomiální funkce, která má spojitě derivace do určitého řádu. Kubický spline je splinem třetího řádu, tj. má spojitou nultou a první derivaci a jeho druhá derivace existuje všude, kromě původních bodů dat (zn. C^1).

Řekneme, že funkce $\varphi: [x_0, x_n] \rightarrow R$ je kubický spline, jestliže

- i. $\varphi \in C^1[x_0, x_n]$,
- ii. $\varphi(x_i, x_{i+1})$ je polynom stupně nejvýše 3.

Říkáme, že φ je kubický interpolační spline k f v bodech x_0, \dots, x_n , jestliže jsou navíc pro $i = 0, 1, \dots, n$ splněny podmínky:

- i. $f(x_i) = \varphi(x_i)$,
- ii. $f'(x_i) = \varphi'(x_i)$.

Restrikci φ na $[x_i, x_{i+1}]$ označíme φ_i . Tuto funkci lze psát ve tvaru

$$\varphi_i(x) = a_i + b_i(x - x_i) + c_i(x - x_i)^2 + d_i(x - x_i)^3. \quad (6)$$

Funkce φ je tedy určena celkem $4n$ parametry. Podmínky z definice kubického interpolačního splinu nám však dávají jen $4n - 2$ podmínek. Dá se tedy očekávat, že budeme muset ještě dva parametry zvolit. Nejčastěji se používají okrajové podmínky v krajních bodech x_0 a x_n , a to tří typů:

- a) $\varphi'(x_0) = f'_0, \varphi'(x_n) = f'_n$;
- b) $\varphi''(x_0) = f''_0, \varphi''(x_n) = f''_n$;
- c) $\varphi''(x_0) = 0, \varphi''(x_n) = 0$.

Přidáním kterékoli dvojice podmínek z a) b) c) nám zajistí jednoznačné definování splinu. Kubický interpolační spline určený podmínkou c) se nazývá *přirozený spline*.

4.3.1.1 Konstrukce kubického splinu

Budeme konstruovat spline určený podmínkou b) (jíž je podmínka c) speciálním případem). Předpokládejme nejprve, že již známe čísla $M_i = \varphi''(x_i)$, tzv. momenty splinu. Funkce φ'' je spojitá a po částech lineární, neboť každá φ''_i je druhou derivací polynomu třetího stupně, tedy polynomem prvního stupně. Označíme-li $h_i = x_{i+1} - x_i$, potom pro $x \in [x_i, x_{i+1}]$ máme

$$\begin{aligned}\varphi_i''(x) &= M_i + (M_{i+1} - M_i) \frac{x - x_i}{x_{i+1} - x_i} \\ &= M_i \frac{x_{i+1} - x}{h_i} + M_{i+1} \frac{x - x_i}{h_i}.\end{aligned}\quad (7)$$

Integrováním dostaneme φ_i' a φ_i :

$$\varphi_i'(x) = -M_i \frac{(x_{i+1} - x)^2}{2h_i} + M_{i+1} \frac{(x - x_i)^2}{2h_i} + A_i \quad (8)$$

$$\varphi_i(x) = M_i \frac{(x_{i+1} - x)^3}{6h_i} + M_{i+1} \frac{(x - x_i)^3}{6h_i} + A_i(x - x_i) + B_i \quad (9)$$

Pomocí (známých) hodnot $\varphi_i(x_i) = f(x_i)$, $\varphi_i(x_{i+1}) = f(x_{i+1})$ určíme konstanty A_i a B_i :

$$f(x_i) = \frac{1}{6}M_i h_i^2 + B_i \Rightarrow B_i = f(x_i) - \frac{1}{6}M_i h_i^2 \quad (10)$$

$$f(x_{i+1}) = \frac{1}{6}M_{i+1} h_i^2 + A_i h_i + B_i \Rightarrow A_i \quad (11)$$

$$= \frac{f(x_{i+1}) - f(x_i)}{h_i} - \frac{h_i}{6}(M_{i+1} - M_i)$$

Zbývá určit hodnoty momentů: M_0 a M_n máme zadané, ostatní vypočteme ze spojitosti první derivace (derivací funkce φ_i v bodech x_i, x_{i+1} se rozumí příslušná jednostranná derivace).

$$\begin{aligned}\varphi'_{i-1}(x_i; -) &= \frac{1}{2}M_i h_{i-1} + A_{i-1} \\ &= \frac{1}{2}M_i h_{i-1} + \frac{f(x_i) - f(x_{i-1})}{h_{i-1}} - \frac{h_{i-1}}{6}(M_i - M_{i-1})\end{aligned}\quad (12)$$

$$\begin{aligned}\varphi'_i(x_i; +) &= -\frac{1}{2}M_i h_i + A_i \\ &= -\frac{1}{2}M_i h_i + \frac{f(x_{i+1}) - f(x_i)}{h_i} - \frac{h_i}{6}(M_{i+1} - M_i)\end{aligned}\quad (13)$$

Z rovnosti obou derivací (12) a (13) dostaneme po úpravách:

$$\begin{aligned}
M_{i-1} \frac{h_{i-1}}{6} + M_i \left(\frac{h_{i-1}}{3} + \frac{h_i}{3} \right) + M_{i+1} \frac{h_i}{6} \\
= \frac{f(x_{i+1}) - f(x_i)}{h_i} - \frac{f(x_i) - f(x_{i-1}))}{h_{i-1}}
\end{aligned} \tag{14}$$

Označíme-li $\lambda_i = \frac{h_{i-1}}{h_{i-1}+h_i}$, $\mu_i = 1 - \lambda_i = \frac{h_i}{h_{i-1}+h_i}$, lze rovnici přepsat ve tvaru

$$\lambda_i M_{i-1} + 2M_i + \mu_i M_{i+1} = g_i \tag{15}$$

kde g_i je pravá strana původní rovnice, vynásobená výrazem $\frac{6}{h_{i-1}+h_i}$. Dostáváme soustavu:

$$\begin{aligned}
2M_1 + \mu_1 M_2 &= g_1 - \lambda_1 f_0'' \\
\lambda_2 M_1 + 2M_2 + \mu_2 M_3 &= g_2 \\
\lambda_3 M_2 + 2M_3 + \mu_3 M_4 &= g_3 \\
&\vdots \\
\lambda_{n-1} M_{n-2} + 2M_{n-1} &= g_{n-1} - \mu_{n-1} f_n''
\end{aligned} \tag{16}$$

Všimněme si, že prvky na diagonále matice soustavy jsou vždy 2, zatímco součet všech ostatních prvků v libovolném řádku je mezi 0 a 1. Matice je tedy ostře diagonálně dominantní, a tedy i regulární (tj. existuje jednoznačné řešení) (Feistauer & Kučera, 2014).

Pokud bychom se rozhodli rovnici kubického splinu sestavit pro každý interval mezi dvěma body samostatně, lze soustavu rovnic sestavit přímo. Např. pro interval $[x_1, x_2]$ by soustava vypadala následovně:

$$\begin{aligned}
\varphi(x_1) &= a + bx_1 + cx_1^2 + dx_1^3 \\
\varphi'(x_1) &= b + 2cx_1 + 3dx_1^2 \\
\varphi(x_2) &= a + bx_2 + cx_2^2 + dx_2^3 \\
\varphi'(x_2) &= b + 2cx_2 + 3dx_2^2.
\end{aligned} \tag{17}$$

Jedná se o regulární soustavu 4 rovnic o 4 neznámých, jejímž vyřešením jednoznačně nalezneme všechny potřebné koeficienty a, b, c, d kubického splinu. Lze dokázat, že tato soustava je regulární, což znamená, že má právě jedno jednoznačné

řešení. Všimněme si, že následující interval $[x_2, x_3]$ by opět vyžadoval použití a úpravy posledních dvou rovnic ze soustavy (17). Jelikož jsou dnes v drtivé většině případů takovéto úlohy řešeny strojově, jakákoli nutnost duplicitního použití dat (v tomto případě rovnic) je neefektivní a nežádoucí. Proto je v případě softwarového řešení lepší použít jednu komplexní soustavu rovnic, jak jsme si ukázali výše (16).

4.3.2 Lagrangeova metoda – konstrukce interpolačního polynomu

Interpolační polynom je první z doposud zmíněných metod, která prokládá všechny naměřené hodnoty jediným spojitým polynomem na celém sledovaném intervalu, nikoliv pouze po částech mezi jednotlivými hodnotami. Platí pro něj:

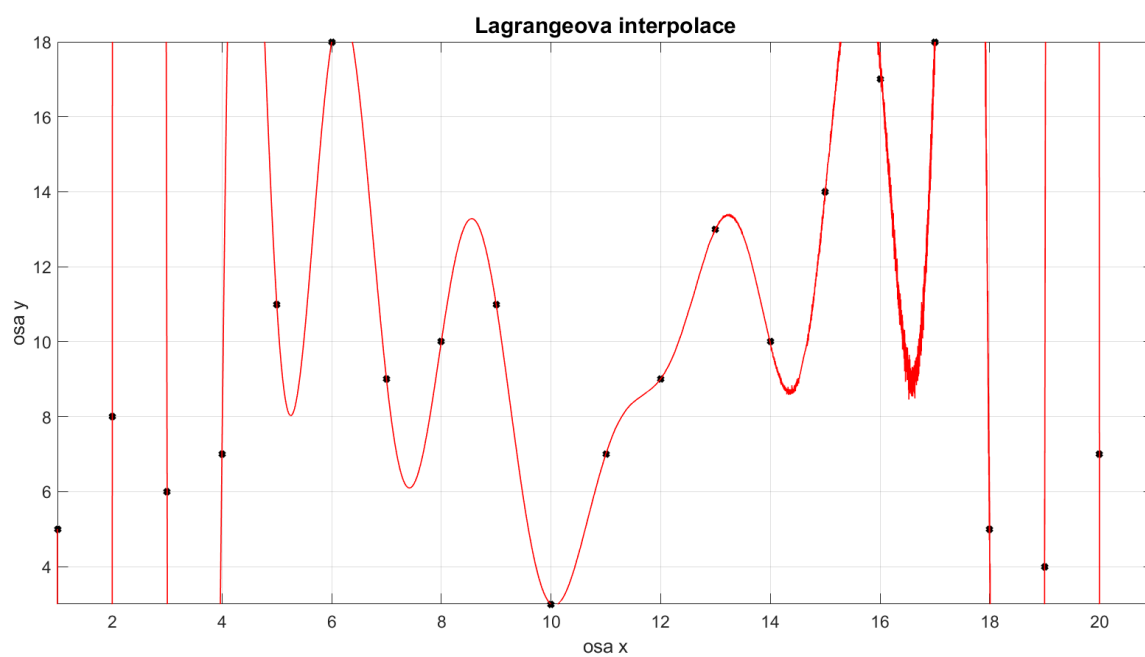
$$P(x) = \sum_{i=0}^n l_i(x)f(x_i), \quad (18)$$

kde

$$l_i(x) = \frac{\prod_{j=0, j \neq i}^n (x - x_j)}{\prod_{j=0, j \neq i}^n (x_i - x_j)}. \quad (19)$$

Opět tak získáváme spojitou hladkou funkci, která přesně prochází všemi naměřenými body. Již na poměrně malém počtu naměřených hodnot se však začne ukazovat, že obzvláště u bodů na krajích sledovaného intervalu, získávají funkční hodnoty interpolačního polynomu velmi vysokou amplitudu, kterou se významně vzdalují oblasti sledovaných dat. I když tak prochází všemi naměřenými body přesně, doplnění funkčních hodnot na intervalech mezi nimi není tímto způsobem téměř vůbec reprezentativní.

Obrázek 1: Graf Lagrangeovy interpolace



Zdroj: vlastní zpracování

5 Fourierovy řady

Fourierovy řady slouží k aproximaci periodických funkcí, kterými mohou být popsány opakující se děje, nebo funkcí na uzavřeném intervalu, který tak v podstatě vyhraní jednu periodu těchto funkcí. Uvažíme-li nejjednodušší periodické funkce $\sin(nx)$ a $\cos(nx)$, mohli bychom 2π -periodickou funkci přiblížit pomocí trigonometrického polynomu

$$T_n(x) = a_0 + \sum_{k=1}^n (a_k \cos(kx) + b_k \sin(kx)). \quad (20)$$

Nechť $\{\varphi_n\}$ je nekonečná ortogonální posloupnost na intervalu $\langle a, b \rangle$, $\{c_n\}$ posloupnost reálných čísel. Nechť řada

$$\sum_{n=1}^{\infty} c_n \varphi_n \quad (21)$$

stejněměrně konverguje k funkci f na intervalu $\langle a, b \rangle$. Potom pro každé n platí:

$$c_n = \frac{(f, \varphi_n)}{(\varphi_n, \varphi_n)} = \frac{(f, \varphi_n)}{\|\varphi_n\|^2}, \quad (22)$$

kde (f, φ_n) je skalární součin f a $\{\varphi_n\}$, tedy integrál jejich součinu přes zkoumaný interval $\int_a^b (f \cdot \varphi_n) dx$ a $\|\varphi_n\|$ je Eukleidovská norma posloupnosti $\{\varphi_n\}$.

Je-li posloupnost $\{\varphi_n\}$ ortonormální, pak se tento vztah zjednoduší na $c_n = (f, \varphi_n)$. Vhodnou volbu takové ortonormální posloupnosti představují například funkce $\sin(nx)$ a $\cos(nx)$ s vhodně zvolenou periodou, které jsou základem pro nejtýpější formu Fourierovy řady (20).

Nechť $\{\varphi_n\}$ je ortogonální posloupnost na intervalu $\langle a, b \rangle$, f je integrovatelná na intervalu $\langle a, b \rangle$. Pak čísla c_n ze vztahu (22) nazýváme *Fourierovy koeficienty funkce f vzhledem k posloupnosti $\{\varphi_n\}$* a řadu (21) *Fourierovou řadou funkce f vzhledem k posloupnosti $\{\varphi_n\}$* (Krisl, 2006).

Výhodou Fourierových řad je, že jsou nekonečně hladké. Mají tedy spojitě derivace všech (nekonečně mnoha) řádů. Další jejich výhodou je také to, že řada (21) se postupně přibližuje k funkci, kterou reprezentuje, proto lze uvažovat konečnou řadu jako hladkou funkci blízkou funkci původní. Tu lze navíc snadno uložit pomocí jejich

koeficientů c_n . Tímto způsobem lze ukládat spojité funkce s velmi malou chybou pomocí konečné posloupnosti $\{c_n\}$, což je výhodné například pro ukládání zvuku do digitální podoby. Fourierovy řady mohou kromě interpolace sloužit také k takzvané aproximaci, kdy chceme hodnoty zadané funkce napodobit jinou funkcí, která bude mít nějaké pro nás přínosné vlastnosti, jako například hladkost, omezenost nebo uložitelnost pomocí menší datové sady, jako jsou právě koeficienty c_n .

6 Regrese

Regresí hodnot funkce (v našem případě diskrétní) rozumíme takovou jejich aproximaci jinou (vhodnou) funkcí, která se jim podle předem zvolené metriky nejvíce blíží. Na rozdíl od interpolace tedy nehledáme funkci, která přímo prochází známými body.

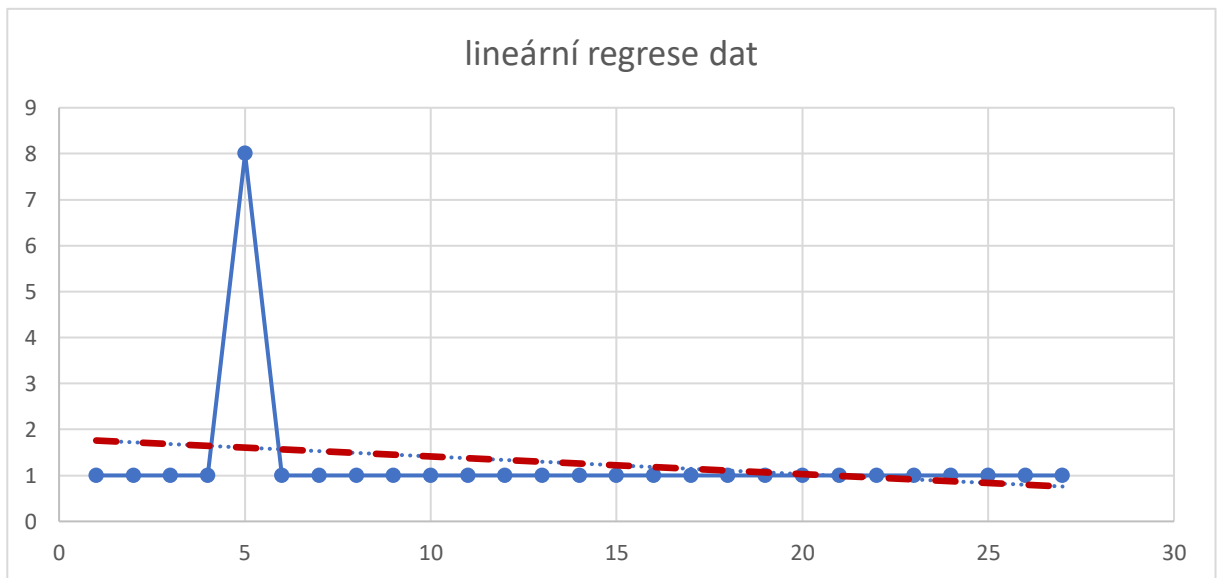
Nejčastěji používanou regresní křivkou je, kvůli své jednoduchosti, přímka

$$f(x) = a \cdot x + b. \quad (23)$$

Obecně lze však zvolit i jiné než lineární regresní funkce, jejichž nejlepší polohu vůči prokládaným hodnotám pak budeme hledat.

Nejpoužívanější metrikou pro určení optimální polohy a vzdálenosti funkce od původních bodů je metoda nejmenších čtverců. O té si řekneme více v kapitole *Měření chyby*.

Obrázek 2: Lineární regrese – příklad vlivu jednoho extrému na jinak konzistentní sadu dat



Zdroj: vlastní zpracování

Na obrázku 2 můžeme vidět odolnost této metody proti jednomu extrémnímu výkyvu v pozorované sadě dat. Tento ojedinělý výkyv samozřejmě naši regresní přímku ovlivní, ale stále zůstane reprezentativní.

6.1 Požadavky na regresní křivku

Chceme-li reprezentovat naměřená data regresní křivkou libovolného typu, typicky na ni klademe jisté požadavky, jejichž splnění nám umožní provádět některé analytické operace.

1. Spojitost – spojitost funkce nám umožňuje získat odhadované hodnoty v libovolném časovém bodě. Tato podmínka je splněna automaticky, je-li funkce hladká.
2. Hladkost – spojitost všech derivací umožňuje např. využít gradientních metod.
3. Jednoduchost – hledaný odhad hodnoty v bodě x_0 získáme tak, že do předpisu naší regresní křivky dosadíme $x = x_0$ a dopočteme funkční hodnotu. Jednoduchost tohoto výpočtu přispívá k rychlosti a všeobecné efektivitě procesu. Navíc je taková křivka pro člověka snadno čitelná.
4. Blížkost datům – samozřejmým předpokladem pro regresní křivku je také to, že věrně reprezentuje aproximovaná data. Čím bližší je křivka naměřeným hodnotám (resp. čím menší je chyba), tím spolehlivější odhad budoucího chování nám poskytne. Více v kapitole *Měření chyby*.

Nejsložitějším úkolem při konstrukci regresní funkce je volba vhodného poměru jednoduchosti a efektivity. Mezi těmito totiž zpravidla platí nepřímá úměra.

Jedná se o statistickou metodu, která může pracovat s různými typy dat. V některých případech se může ukázat, že lineární funkce není výrazně výhodnější regresí z hlediska naměřené chyby, než konstantní hodnota. V takových případech nelze sklon regresní přímky vnímat jako trend dat. Tato problematika není tématem mé práce, na její otázky odpovídají například metody χ^2 .

7 Extrapolace

Extrapolace je aproximační metoda, která odhaduje hodnoty funkce mimo interval známých hodnot. Jde v podstatě o opak interpolace, která hledá hodnoty funkce uvnitř tohoto intervalu (Pikal, 2020). Jedním způsobem, jak této metody využít, je predikce vývoje zkoumané veličiny na základě znalosti dat z minulosti. Zde je důležité si uvědomit, že přesnost takového odhadu se s rostoucí vzdáleností od známých hodnot v čase poměrně rychle snižuje. Je proto vhodné pokoušet se odhadovat pouze chování dané veličiny v nejbližší době po poslední známé hodnotě. Zároveň i zde platí obecné statistické pravidlo, že čím více dat již máme vlivem pozorování a měření k dispozici, tím přesnější analýzu a odhad zkoumané veličiny jsme schopni učinit. Úspěšnost odhadu samozřejmě stojí i na tom, že zkoumaná veličina se chová podobným způsobem v době měření, jako v době, kterou budeme odhadovat. Pokud se jedná o chaotický jev nebo dojde k výrazné změně v systému, nemusí být předchozí data při odhadu použitelná.

Dalším praktickým využitím této metody může být například doplnění krajních hodnot v již dříve zmíněném kubickém splinu. Jestliže je kubický spline po částech tvořen polynomy 3. stupně, které na sebe hladce navazují (se spojitou nultou i první derivací), pak jeho druhá derivace musí být po částech lineární. V případě extrapolovaného splinu se hodnota druhé derivace v prvním bodě splinu x_0 bude nacházet na polopřímce tvořené body x_2 a x_1 , v posledním bodě na polopřímce x_{n-2} , x_{n-1} (Pikal, 2020).

7.1 Využití extrapolace při konstrukci kubického splinu

Jestliže body $[x_0, m_0]$, $[x_1, m_1]$, $[x_2, m_2]$ leží na jedné přímce, můžeme pro nalezení m_0 použít rovnici přímky: $y = kx + q$. Dosazením druhého a třetího bodu dostaneme soustavu rovnic:

$$\begin{aligned}m_1 &= kx_1 + q \\m_2 &= kx_2 + q\end{aligned}\tag{24}$$

Z první rovnice vyjádříme q a k získáme odečtením první rovnice od druhé:

$$q = m_1 - kx_1, k = \frac{m_2 - m_1}{x_2 - x_1}\tag{25}$$

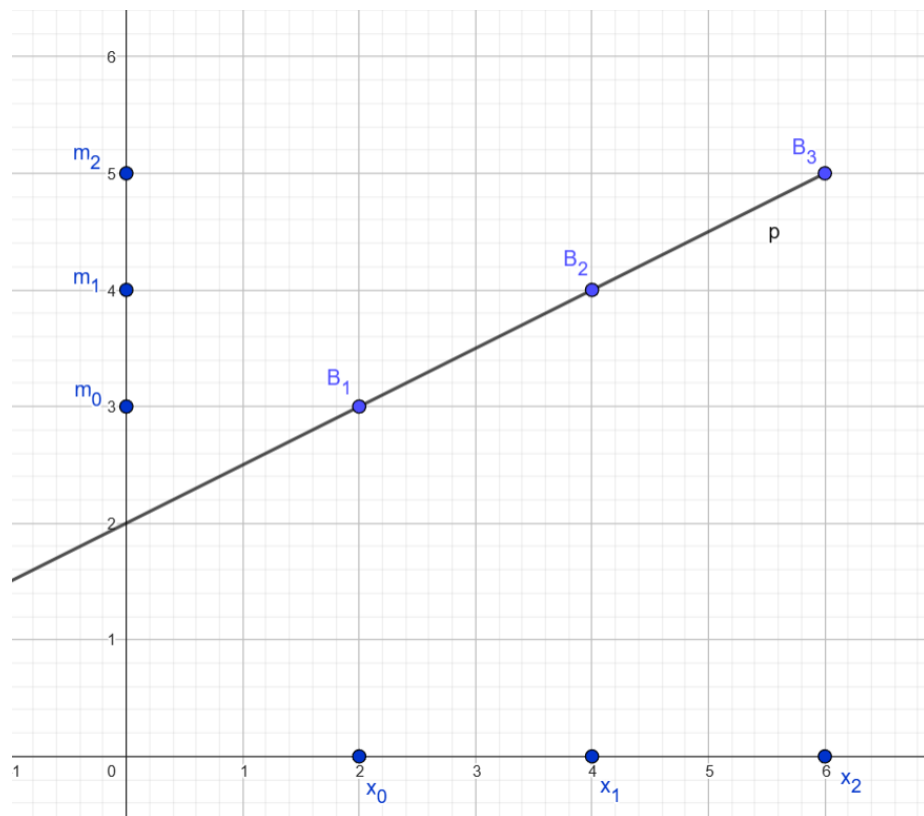
Hodnotu druhé derivace v bodě x_0 dostaneme, když do $m_0 = kx_0 + q$ dosadíme q a k z (25). Po úpravě by vyjádření vypadalo takto:

$$m_0 = m_1 + \frac{(x_1 - x_0)(m_2 - m_1)}{x_2 - x_1} \quad (26)$$

Stejným postupem s body $[x_{n-2}, m_{n-2}]$, $[x_{n-1}, m_{n-1}]$, $[x_n, m_n]$, vypočítáme hodnotu druhé derivace v posledním bodě splinu:

$$m_n = m_{n-1} + \frac{(x_n - x_{n-1})(m_{n-1} - m_{n-2})}{x_{n-1} - x_{n-2}} \quad (27)$$

Obrázek 3: Souřadnice bodů na přímce



Zdroj: vlastní zpracování

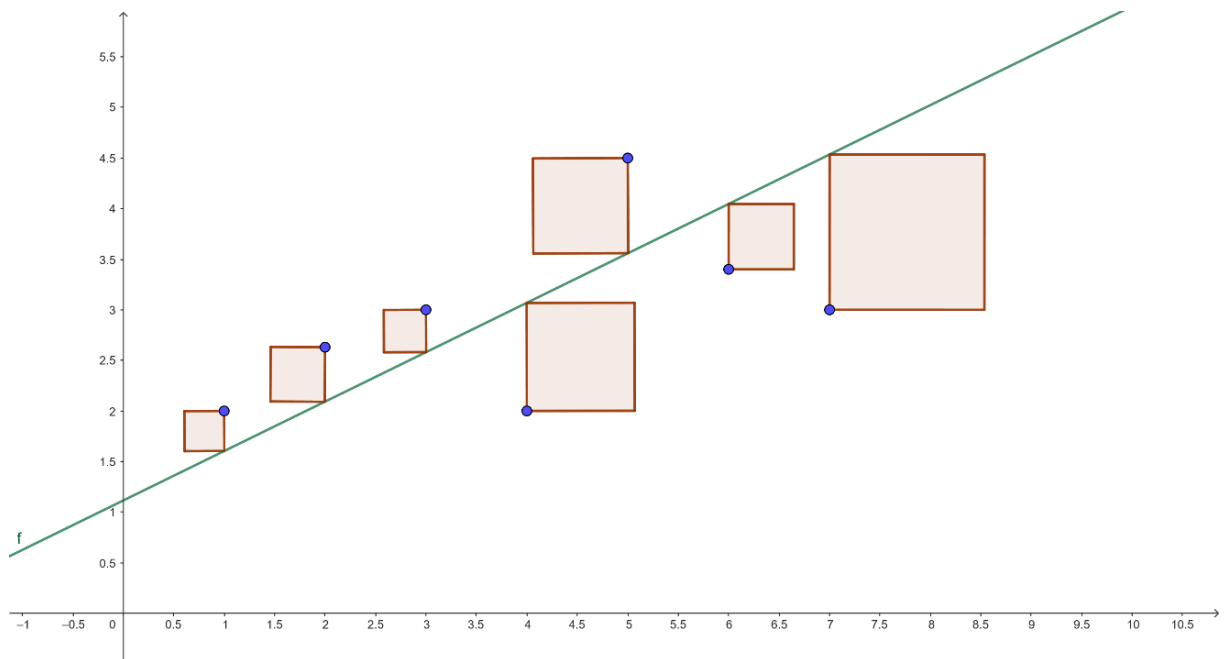
8 Měření chyby

8.1 Metoda nejmenších čtverců

Metoda nejmenších čtverců slouží k nalezení optimálního tvaru a umístění námi zvolené funkce ve smyslu její minimální celkové vzdálenosti od naměřených bodů.

Regresní křivky nabízí možnost reprezentovat libovolně velkou množinu dat hladkou, přitom velmi jednoduše (řádově jednotkami parametrů) popsatelnou funkcí, jejíž typ si navíc můžeme sami zvolit. Také přináší velikou výhodu ve své poměrné odolnosti proti malému množství odlehlých dat (způsobených například chybou měření či málo pravděpodobnou náhodou), která jsou vzdálená od jinak stabilního trendu. Při použití interpolační křivky by takový extrém velmi výrazně ovlivnil její tvar a mohl tak znehodnotit její reprezentaci skutečnosti.

Obrázek 4: Metoda nejmenších čtverců



Zdroj: vlastní zpracování

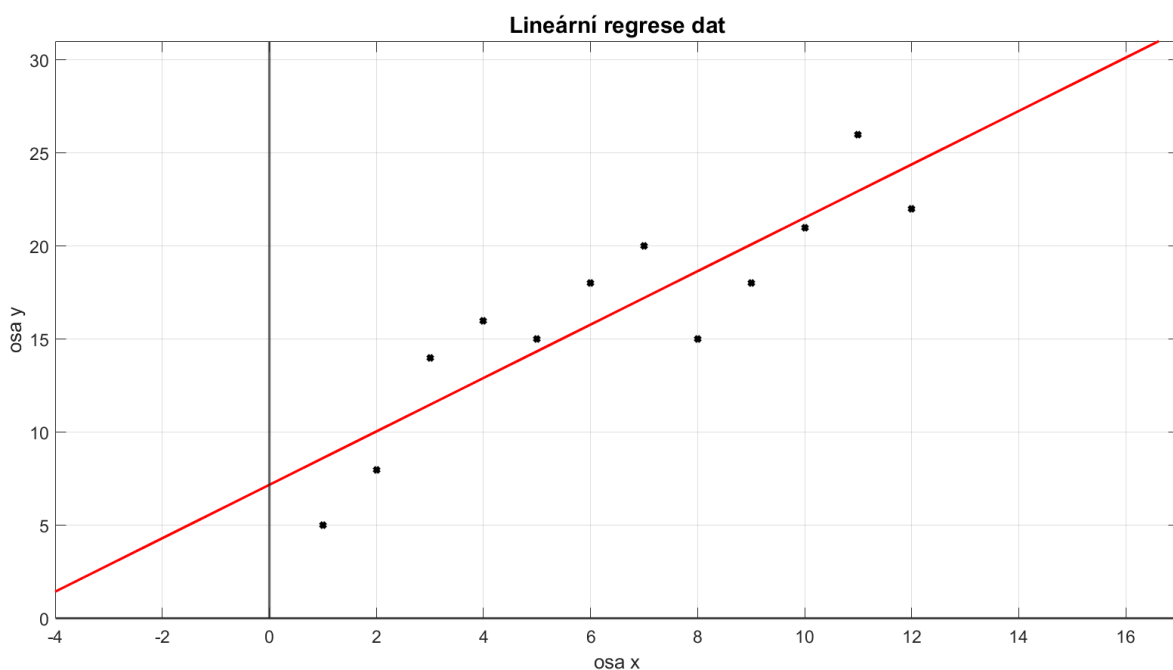
Metoda nejmenších čtverců definuje celkovou chybu jako součet čtverců rozdílů naměřených hodnot y_i a regresní křivky $f(x)$ (28).

$$E = \sum_{i=0}^n |f(x_i) - y_i|^2 \quad (28)$$

Pro získání minima této funkce (tedy minimální chyby) pro regresní přímku vzhledem k jejím parametrům a a b je potřeba najít stacionární bod x_s . Pro ten platí $E'(x_s) = 0$. Jeho nalezením získáme rovnice (29), z nichž parametry a a b vypočteme.

$$\begin{aligned} a \cdot \sum_{i=1}^n x_i^2 + b \cdot \sum_{i=1}^n x_i &= \sum_{i=1}^n x_i y_i, \\ a \cdot \sum_{i=1}^n x_i + b \cdot n &= \sum_{i=1}^n y_i. \end{aligned} \quad (29)$$

Obrázek 5: Graf lineární regrese dat metodou nejmenších čtverců



Zdroj: vlastní zpracování

8.2 Lebesgueovy prostory

Další možností, jak změřit celkovou chybu aproximace a určit tak optimální polohu a tvar regresní funkce z funkce původní, jsou Lebesgueovy prostory. Jedná se o velmi komplexní téma, které je blíže popsáno například v (Lukeš & Malý, 2002). Pro naše účely se z této kapitoly budeme zabývat pojetím chyby, kterou budeme uvažovat následovně:

$$E_p = \sqrt[p]{\sum_{i=0}^n |e_i|^p}, \quad (30)$$

$$e_i = f(x_i) - a_i. \quad (31)$$

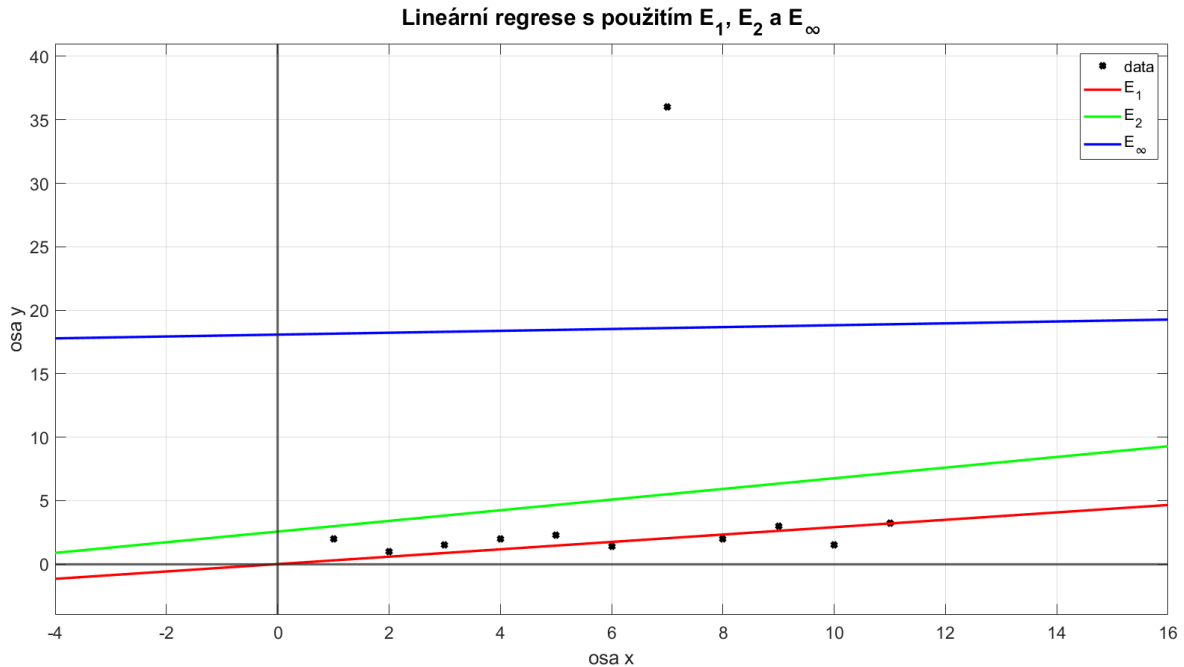
Rovnice (30) a (31) popisují výpočet celkové chyby E_p při aproximaci posloupnosti a_n (tj. naměřených hodnot) funkcí $f(x)$. Opět tedy uvažujeme aproximaci diskrétních hodnot spojitou funkcí.

Parametr p v rovnici (30) může být obecně libovolné kladné číslo, obvykle větší rovno 1. Nejčastěji používanými případy, kterými se i my budeme zabývat, jsou:

- a) E_1 – zde se jedná o prostou sumu vzdáleností našich dvou funkcí v jednotlivých bodech měření započítanou nezávisle na vzájemné poloze, tedy absolutní hodnotu z rozdílu funkčních hodnot. Použití této metody zajistí, že od všech bodů bude aproximační funkce podobně vzdálená a všechny tyto body stejnou měrou ovlivní její tvar i polohu.
- b) E_2 – lze říci, že E_2 je nejuniverzálnější a používá se nejčastěji. Důvodem je především to, že je jakýmsi kompromisem mezi zbylými dvěma zmíněnými a) a c). Jedná se prakticky o metodu nejmenších čtverců, kterou jsme si popsali výše. Práce s druhou mocninou zdůrazňuje „důležitost“ velkých vzdáleností, ale zároveň stále bere v potaz vzdálenosti funkce od všech bodů.
- c) E_∞ – tato metoda v podstatě měří celkovou chybu jako maximum ze vzdáleností naší aproximační funkce od všech bodů. Přestože nekonečno nelze při výpočtech dosazovat za p do rovnice (30), použitím velmi vysokých hodnot dostáváme předpis, který odpovídá hodnotě nazývané E_∞ . Jediná hodnota extrémně vychýlená od ostatních zde velmi výrazně ovlivní, jak bude naše funkce vypadat.

Obrázek 6 ukazuje rozdíl při použití E_1 , E_2 a E_∞ při lineární regresi na stejné sadě dat s podobným trendem a jedním extrémem. Zatímco E_1 je takovým extrémem ovlivněno zdaleka nejméně a E_2 stále jen málo, E_∞ se posune na půl cesty od původních dat.

Obrázek 6: Lineární regrese s použitím E_p



Zdroj: vlastní zpracování

8.3 Sobolevovy prostory

Dalším silným nástrojem matematické analýzy funkcí jsou Sobolevovy prostory. Ty se mimo jiné využívají také v teorii distribuce nebo při řešení parciálních diferenciálních rovnic. Ve spojitém pojetí představují velice komplexní oblast matematiky, která by vydala na samostatné práce, jako například (Leoni, 2017).

Jelikož se v této práci zaměřujeme na analýzu diskretních dat měřených v pravidelných intervalech, vystačíme si s aplikací Sobolevových prostorů v diskretním pojetí. V takovém je lze poměrně snadno aplikovat na chápání změny hodnot zkoumaného systému jako důsledek změny jeho energie (každá změna hodnoty je způsobena přidáním energie do systému). Derivace naší diskretní funkce ve všech bodech pak ukazuje, kolik energie v sobě systém má, tj. jak moc je klidný, či proměnlivý. Opět budeme uvažovat tři základní měření chyby: $E_1(32)$, $E_2(33)$ a $E_\infty(34)$.

$$E_1 = \sum_{i=0}^n |f'(x_i)| \quad (32)$$

$$E_2 = \sqrt{\sum_{i=0}^n |f'(x_i)|^2} \quad (33)$$

$$E_\infty = \sqrt{\sum_{i=0}^n |f'(x_i)|^\infty} \quad (34)$$

Různé formy energie jsou významné při modelování procesů a nahrazování složitějších funkcí jednoduššími, zejména tak, aby byl rozdíl energií původní funkce a její aproximace malý. Alternativně bychom při vyhodnocení vhodnosti aproximační metody mohli srovnávat i Sobolevovskou normu, tedy energii rozdílu obou funkcí. Takové pojetí ale již přesahuje rozsah této práce.

9 Aplikace interpolačních a regresních metod na měnové kurzy

Pro Českou republiku, jakožto členskou zemi Evropské unie s ekonomikou silně zaměřenou na import a export, je jedním z klíčových faktorů měnový kurz. Pojďme se proto při aplikaci výše zmíněných metod zaměřit na vývoj kurzu EUR/CZK devizového trhu. Ten udává sílu eura vůči koruně, tedy za kolik Kč je možné koupit 1 €.

V přílohách A až C jsou k dispozici ceny eura (v českých korunách) zaznamenávané po řadě měsíčně, čtvrtletně a ročně od roku 1999 do současnosti převzaté z České národní banky (*Kurzy devizového trhu – měsíční průměry - Česká národní banka, b.r.*).

9.1 Interpolace

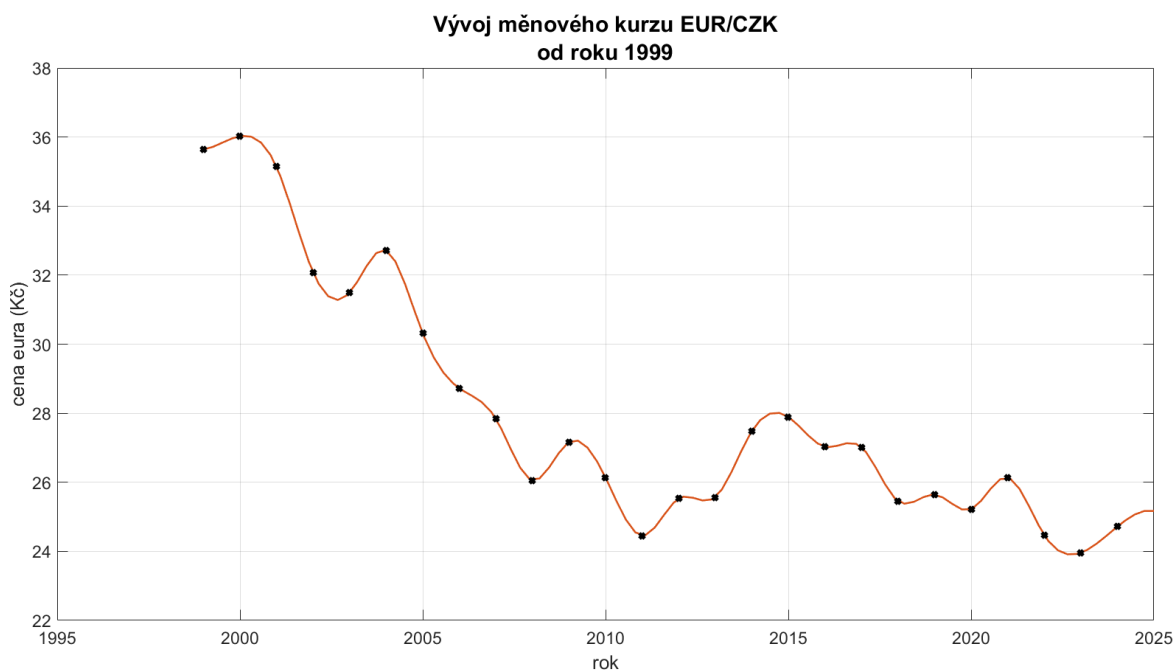
Na obrázku 7 můžeme vidět graf vývoje kurzu měřený na začátku každého roku interpolovaný kubickým splinem, jehož předpisy pro jednotlivé intervaly jsou přiložené v příloze D.¹ Porovnáme-li tento spline s hodnotami skutečně naměřenými ve čtvrtletích, dostaneme průměrnou chybu E_1 přibližně 0,41 €. Reálně naměřená data po měsících se od splinu průměrně vzdalují dokonce o 0,44 €. To znamená, že při zhuštění dat skutečně dochází k očekávanému zpřesnění informace, resp. použitím méně dat vzrůstá vzdálenost odhadu od skutečných hodnot ve vynechaných bodech.

Potvrzuje se tím ve statistice obecně platné tvrzení, že každá informace navíc nám umožňuje přesnější závěry (v našem případě odhad). Otázkou ke zvážení ovšem zůstává, o kolik přesnější odhad jsme získáním dalších informací (zhuštěním naměřených dat) schopni udělat a jak náročné je takové zhuštění provést. V tomto konkrétním případě jsme získali zpřesnění 0,03 € (tedy přibližně 0,11 % z průměrné hodnoty zkoumané veličiny) za cenu třikrát častějšího měření.

Je důležité si uvědomit, že každé statistické měření a sběr dat představuje jednak další náklady (jak časové, tak finanční) a jednak je samo o sobě zásahem do systému zkoumané veličiny, čímž ji může ovlivňovat. Provádět příliš častá měření tedy není žádoucí.

¹ Viditelné nepřesnosti grafu (ostré hrany) jsou způsobeny technologickou nedokonalostí MATLABu. Předpisy funkcí jsou po částech kubické polynomy, graf je tedy skutečně hladký.

Obrázek 7: Kubický spline – vývoj kurzu EUR/CZK



Zdroj: vlastní zpracování

Přestože manuální konstrukce kubického splinu, resp. sestavení soustavy (16) a její řešení je poměrně náročné a vyžaduje pokročilé matematické znalosti, jeho aplikace pomocí softwaru² je s použitím správných funkcí velice jednoduchá jak pro uživatele, tak z výpočetního hlediska.

Není tedy třeba metodu soudit podle její složitosti a můžeme se zabývat pouze přesností aproximace, kterou nám poskytuje.

Tabulka 1: Kubický spline – MATLAB funkce

```
% načte data z excelu
T = readtable("eur_to_czk_all_time.xlsx");
A = table2array(T);
Y = A(:,2);
X = 1999:1:1999+length(Y)-1;

% vytvoří kubický interpolační spline
spline_interp = spline(X, Y);

% předpis kubického spline
coefficients = spline_interp.coefs;

% vypíše předpisy pro jednotlivé intervaly
```

² MATLAB má v sobě implementované funkce pro práci se splinem použité v tabulce 1.


```

for i = 1:length(coefficients)
    fprintf('Pro interval [%d, %d]:\n', X(i), X(i+1));
    fprintf('s(x) = %.4fx^3 + %.4fx^2 + %.4fx + %.4f\n', coefficients(i,1),
coefficients(i,2), coefficients(i,3), coefficients(i,4));
end

%% vykreslí
plot(X, Y, 'kx', linspace(1999,1999+length(Y)), ppval(spline_interp,
linspace(1999,1999+length(Y))), '-', 'linewidth',1.5);
hold on
plot(X, Y, 'kx', 'linewidth',3)
grid on
xlabel('rok');
ylabel('cena eura (Kč)');
title("Vývoj měnového kurzu EUR/CZK" + newline + "od roku 1999");
hold on

```

Zdroj: vlastní zpracování

Podíváme-li se na systém z hlediska energie Sobolevových prostorů, dostaneme hodnoty uvedené v tabulce 2. Ty ukazují, že velký vliv na hodnotu jeho energie hraje škálování dat, resp. jejich hustota, kterou se rozhodneme použít. Proto je důležité vždy na začátku rozmyslet, jaké hodnoty chceme doplňovat (či později odhadovat) a na základě toho zvolit vhodná vstupní data a vhodné přístupy. Například modelovat měsíční hodnoty na základě ročních měření se tímto ukazuje velmi nepřesné.

Tabulka 2: Energie systému EUR/CZK z hlediska hustoty měření

energie/měření	ročně	čtvrtletně	měsíčně
celková	27,05	42,22	78,13
průměrná	1,04	0,42	0,26
maximální	3,06	2,25	2,08
minimální	0,01	0	0

Zdroj: vlastní zpracování

Dalším způsobem využití interpolace, jak jsme si výše uvedli, je doplnění chybějících hodnot uvnitř intervalu. Pokud by například hodnoty pro rok 2010 z nějakého důvodu nebyly k dispozici (ať už proto, že měření neproběhlo, nebo se údaje nedochovaly), mohli bychom se pokusit je interpolovat ze známých okolních hodnot.

Z výsledné aproximace kubickým splinem dostaneme odhadovanou cenu eura 25,89 Kč. Skutečně naměřená cena v roce 2010 byla 26,14 Kč. Náš odhad je tedy vzdálený od skutečnosti o méně než 1 %.

Pokud bychom ovšem neměli k dispozici ani data z roku 2011, činil by odhad pro rok 2010 27,5 Kč a pro rok 2011 26,6 Kč, což oproti skutečným 26,14 Kč, resp. 24,45 Kč činí rozdíl 5,2 %, resp. 8 %. To nám ukazuje, že odhadovat jednu chybějící hodnotu na dostatečně hustém vzorku je s poměrně vysokou přesností možné, nicméně pokud chybí údajů více, přesnost odhadu, který jsme schopni učinit, velice rychle klesá.

9.2 Lineární regrese

Proložíme-li data regresní přímkou (obrázek 8), ta bude mít (s užitím optimalizace chyby pomocí E_2) předpis:

$$y(x) = -0,427x + 886,871. \quad (35)$$

Dosazením do (35) zjistíme, že odhadem kurzu pro rok 2023 je 1 € = 23,05 Kč. Z přílohy C vidíme, že skutečná hodnota kurzu byla 23,96 Kč za euro, což činí cca 3,8% rozdíl. V tomto případě platí:

$$r = r_0 \pm 5\%, \quad (36)$$

kde r je odhadovaná hodnota a r_0 je skutečná hodnota kurzu. Vidíme, že odhad se nachází v relativně blízkém okolí³ reálné hodnoty, můžeme ho tedy považovat za vhodný.

Použitím optimalizace chyby pomocí E_∞ bychom získali regresní přímkou s předpisem

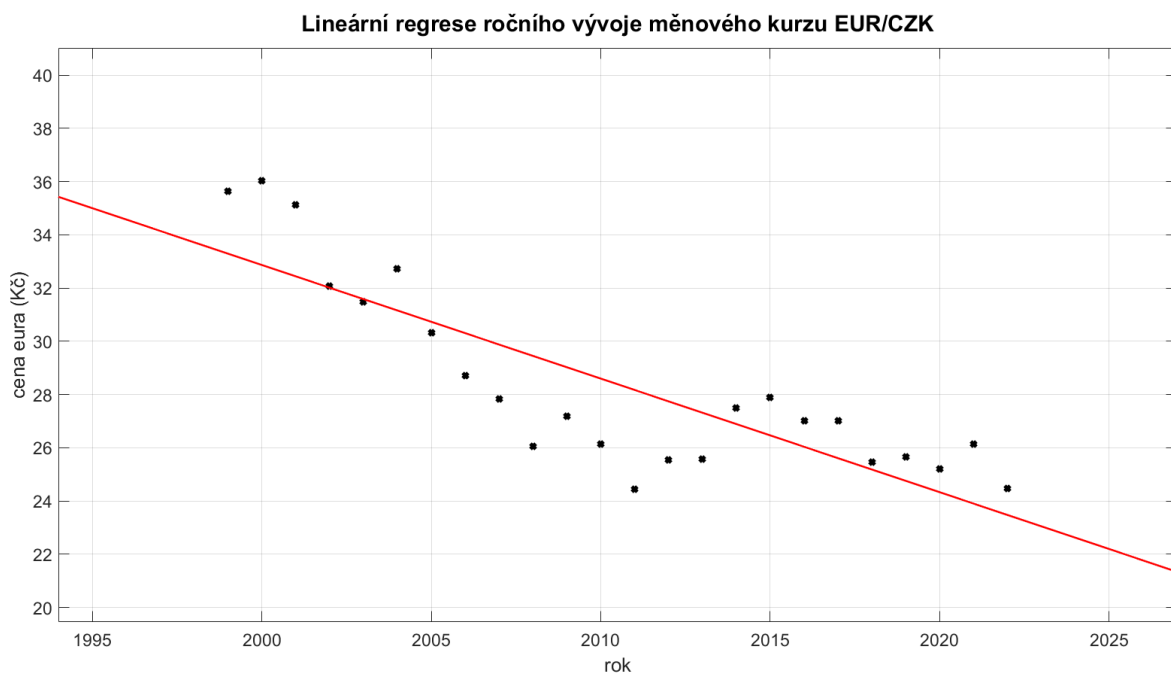
$$y(x) = -0,471x + 974,49. \quad (37)$$

Odhadovaný kurz pro rok 2023 by tak dosazením do (34) byl 21,66 Kč za euro, tedy přibližně 9,6% rozdíl oproti skutečné hodnotě naměřené v daném roce. Vidíme tedy, že odhad (36) již není splněn a tato přímkou není pro odhad budoucího kurzu vhodná.

³ Ve statistice obecně považujeme 5% odchylku za akceptovatelnou.

Je důležité zdůraznit, že odhad lineární regrese obecně ztrácí u většiny typů dat velmi rychle na smysluplnosti. Regresní přímka totiž ze své podstaty umožňuje zachytit pouze lineární trend dat. Pokud data střídavě rostou a klesají, rostou jiným než lineárním tempem, nebo vykazují jiné nepravidelné chování, je aproximace takových dat přímkou krajně nevhodná a je zapotřebí zvážit použití jiné funkce. Volba funkce závisí na nalezení podobnosti dat s běžnými křivkami funkcí. Podobně, jako aproximujeme funkci tečnou, může být lineární regrese vhodná na krátkém časovém úseku, kde se kolísání ani dlouhodobý nelineární trend neprojeví.

Obrázek 8: Lineární regrese (E_2) – vývoj kurzu EUR/CZK



Zdroj: vlastní zpracování

9.3 Polynomiální regrese

Jak jsme si demonstrovali výše, lineární regrese nemusí být pro naše zkoumaná data vždy vhodná. Pojdme se tedy podívat, jak by vypadalo použití polynomu vyššího stupně. I zde lze snadno použít již existující funkce implementované v MATLABu.

Tabulka 3: Regrese polynomem vyššího stupně – MATLAB funkce

```
clc; close all;

%% načtení dat z excelu
T = readtable("eur_to_czk_all_time-do2022.xlsx");
A = table2array(T);

Y = A(:,2)';
X = 1999:1:1999+length(Y)-1;

% Definice funkce E
koef = 2;
E = @(params) nthroot(sum(power(abs(Y - polyval(params,X)),koef)),koef);

% Počáteční odhad parametrů (koeficienty polynomu)
params0 = zeros(1,13);

% Nalezení optimálních parametrů
optimal_params = fminsearch(E, params0);

% Výpis výsledků
format shortE
display(optimal_params);

%% Vykreslení
border = 5;

% body
figure ('Name', 'E_2');
plot(X, Y, 'bx', 'linewidth',3);
grid on;
title('Polynomiální regrese');
xlabel('osa x')
ylabel('osa y')
xlim([min(X)-border max(X)+border])
ylim([min(Y)-border max(Y)+border])
hold on;

% polynom
fplot(@(X) polyval(optimal_params,X), [min(X)-border,max(X)+border], 'r-',
'LineWidth', 2);

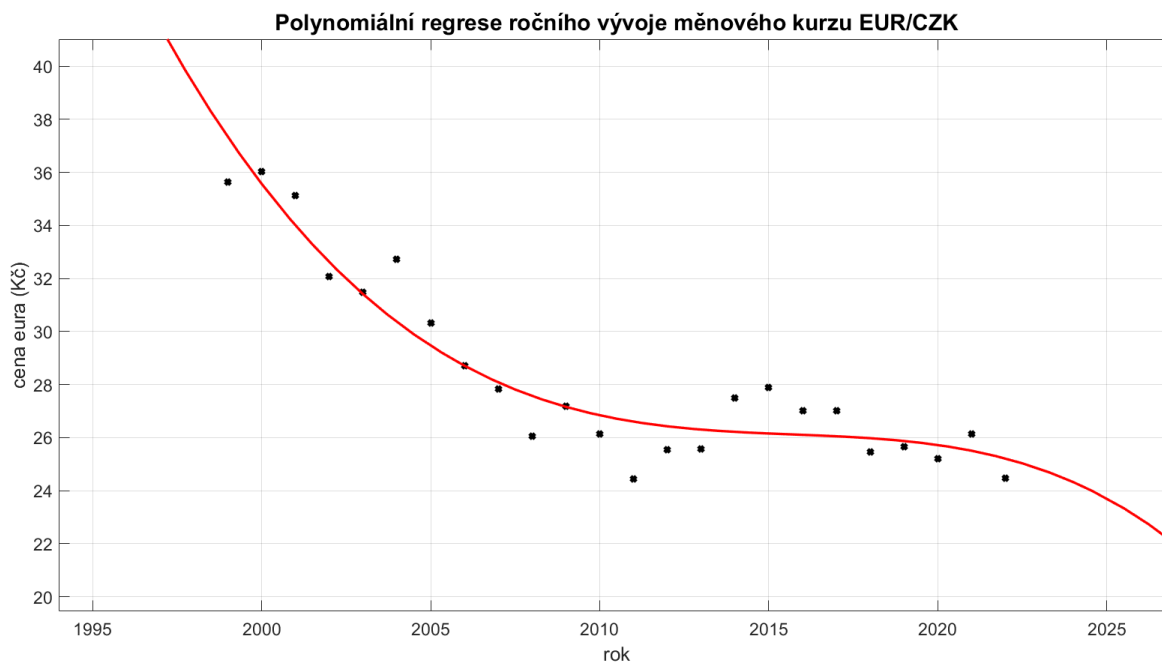
% osy
xline(0,'linewidth',2);
yline(0,'linewidth',2);
hold off;
```

Zdroj: vlastní zpracování

Použitím kódu z tabulky 3 dostaneme předpis polynomu 12. stupně, který se v blízkém okolí zkoumaných dat nechová jako přímka a nabízí tedy přesnější aproximaci. V případě, že nepoužíváme software a počítáme celou situaci ručně, může být práce

s polynomy takového stupně nepřiměřeně náročná. V případě využití dostupných softwarů jde ale stále o reprezentaci jediným vektorem o 13 složkách, jehož zpracování je triviální.

Obrázek 9: Polynomiální regrese (E_2) – vývoj kurzu EUR/CZK



Zdroj: vlastní zpracování

Dosažením do polynomu zobrazeného na obrázku 8 dostaneme odhadovanou cenu eura pro rok 2023 24,82 Kč. To oproti skutečné ceně z přílohy C (23,96 Kč za euro) činí pouze 3,6 %, což opět splňuje rovnici (36) a zároveň je skutečné hodnotě blíže než při předchozím použití regresní přímky.

Přestože je toto přiblížení poměrně malé, připomeňme si opět, že jsme ho docílili pouhou volbou stupně polynomu, což v dobře připraveném počítačovém programu nepředstavuje žádnou práci navíc.

Použitím metody E_∞ na konstrukci polynomu stejného řádu dostaneme odhadovanou cenu 25,95 Kč se vzdáleností od skutečné ceny 8,3 %. I zde se nám tedy potvrzuje, že použití metody E_∞ pro určení chyby při volbě polohy regresního polynomu není vhodnější, než častěji používaná E_2 .

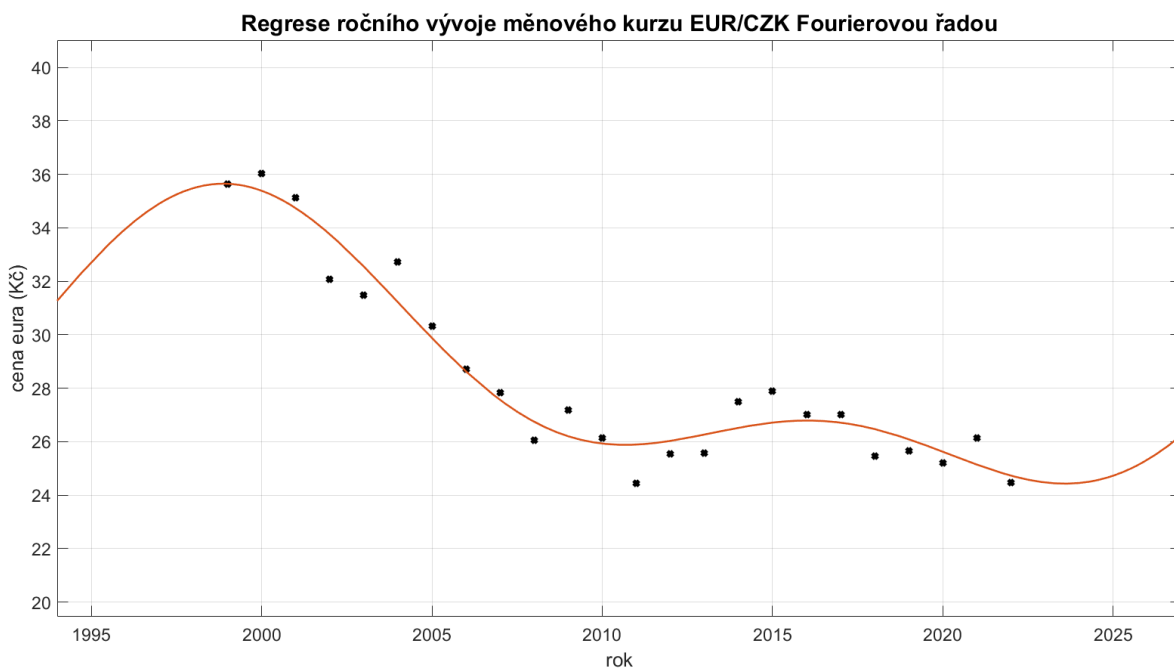
9.4 Regrese Fourierovou řadou

Jak jsme si ukázali v předchozích kapitolách, polynomy nejsou jediný typ funkcí, který se může pro regresi ukázat jako vhodný. Obzvláště u ekonomických funkcí, které často vykazují charakter určitých cyklů, může použití periodických funkcí přinést lepší (přesnější) výsledky. Za pomoci kódu v příloze F lze zjistit, že použitím Fourierovy řady se dvěma harmonickými složkami, jejíž obecná podoba je znázorněna rovnicí (38), dosáhneme také poměrně přesných výsledků. Její konkrétní koeficienty pak zjistíme kódem v příloze G. Zde je důležité zdůraznit, že funkce má více než dva parametry. Není proto jisté, že algoritmus vždy nalezne její optimální polohu, jako je tomu se dvěma parametry u lineární regrese.

Získaný odhad ceny je zde 24,47 Kč, který vykazuje chybu 2,1 %. Opět tedy vidíme vyhovující odhad z hlediska splnění požadavku (36) a poměrně výrazné zpřesnění odhadované informace. Zde už bylo ovšem zapotřebí použití více programových kódů a jejich správná interpretace, což lze považovat za mírné zkomplikování. Stále lze ale konstatovat, že použití této metody je obecně jednoduché, a především snadno opakovatelné.

$$F(x) = a_0 + a_1 \cos(x\omega) + b_1 \sin(x\omega) + a_2 \cos(2x\omega) + b_2 \sin(2x\omega) \quad (38)$$

Obrázek 10: Regrese Fourierovou řadou (E_2) – vývoj kurzu EUR/CZK



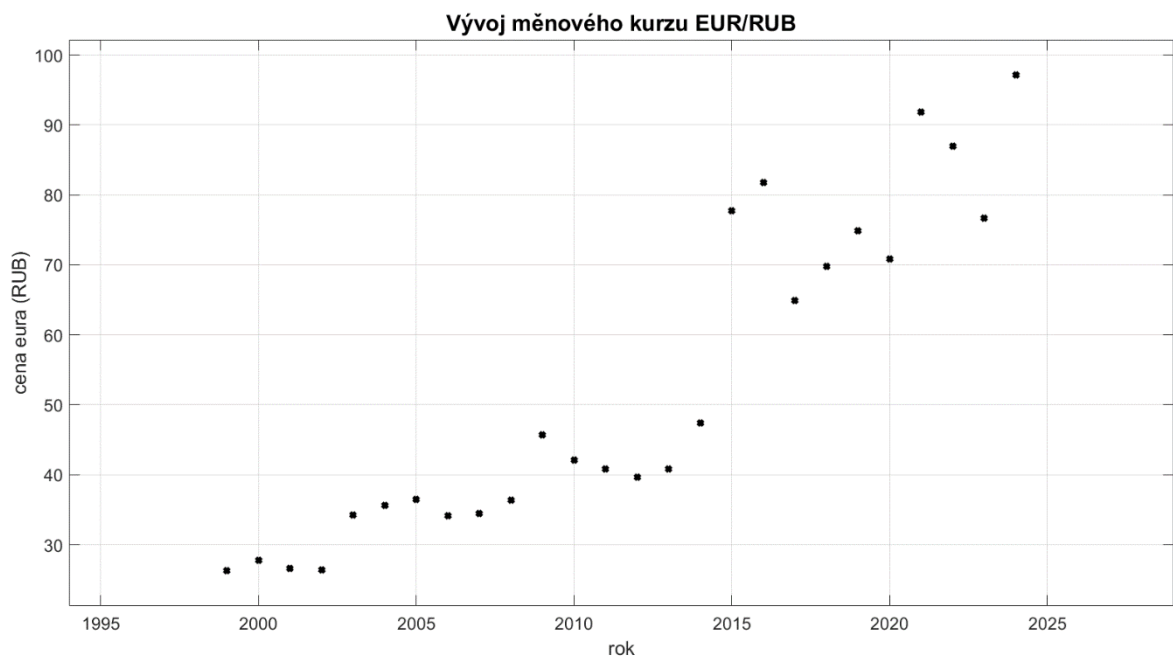
Zdroj: vlastní zpracování

9.5 Regrese v čase neobvyklých událostí

Doposud jsme se věnovali zkoumání dat, která mají poměrně mírný trend změny a dají se tak poměrně přesně zvolenými metodami odhadovat. Pojdme si nyní ukázat, jak se výše popsané regresní metody chovají na „divočejších“ datech, která mají velké rozpětí mezi minimální a maximální hodnotou a mezi těmito hodnotami se vyskytují nekonzistentně.

Příkladem takových dat může být opět měnový kurz, pouze se stačí zaměřit na jiné země, resp. jiné měny. V této kapitole budeme jako základní sadu dat používat měnový kurz devizového trhu EUR/RUB⁴ dostupný na portálu *Investing*, přiložený v příloze E (*EUR RUB Historical Data*, 2024).

Obrázek 11: Vývoj měnového kurzu EUR/RUB



Zdroj: vlastní zpracování

Aplikujeme-li stejné metody, jako v předchozí kapitole na tyto data a pokusíme se z hodnot naměřených v letech 1999 až 2004 odhadnout cenu eura v roce 2005 (skutečná cena byla 36,56 ₣), dostaneme údaje zobrazené v tabulce 4.

⁴ Kurz EUR/RUB udává sílu eura vůči rublu, tedy za kolik rublů je možné koupit 1 ₣.

Tabulka 4: Různé regrese kurzu EUR/RUB pro rok 2005 a jejich chyby

metoda	odhad ceny eura 2005 [P]	chyba [%]
lin. reg. E_2	36,07	1,34
lin. reg. E_∞	35,33	3,36
polynom. reg. E_2	36,16	1,09
polynom. reg. E_∞	35,4	3,17
Fourier. reg.	27,87	23,77

Zdroj: vlastní zpracování

Zatím, kromě použití Fourierovy řady, která se vlivem malého vzorku dat značně vychyluje, nevidíme nic závažně odlišného od předchozích výsledků. Zdá se tedy, že i na kurz EUR/RUB lze tyto odhady bezpečně aplikovat. Pokud ovšem totéž provedeme o 10 let později (tj. pro odhad ceny v roce 2015, kdy skutečná hodnota činila 77,76 P), dostaneme údaje zobrazené v tabulce 5.

Tabulka 5: Různé regrese kurzu EUR/RUB pro rok 2015 a jejich chyby

metoda	odhad ceny eura 2015 [P]	chyba [%]
lin. reg. E_2	46,74	40,0
lin. reg. E_∞	48,49	37,6
polynom. reg. E_2	45,63	41,3
polynom. reg. E_∞	48,74	37,3
Fourier. reg.	49,83	35,9

Zdroj: vlastní zpracování

Jak je možné, že tatáž metoda aplikovaná na stejnou funkci na jedné části zkoumaného intervalu funguje velice přesně a na jiné naprosto selhává? Podívejme se znovu na obrázek 11. Už ze samotných dat je vidět, že v roce 2015 došlo k výrazné změně

chování dat, která oproti původnímu (poměrně stabilnímu) trendu začala nárazově uskakovat.

U modelové matematické funkce samozřejmě takové chování standardně neočekáváme. Je ale potřeba si opět připomenout, že zde zkoumáme data získaná měřeními a pozorováními reálného světa, kde mohou nastávat nečekané události, které vstoupí do systému a velmi výrazně ho ovlivní. Přesně taková událost nastala v roce 2014, kdy proběhla anexie Krymu Ruskou federací. Tato akce byla odsuzována jak Evropskou unií, tak Spojenými státy a vedla k protiruským sankcím, které měly za cíl trestat ruskou agresi a podpořit Ukrajinu. Jelikož Rusko bylo již tehdy významnou součástí Evropské ekonomiky zejména v oblasti energetiky a obchodu, měly tyto sankce obrovský dopad nejen na sílu rublu.

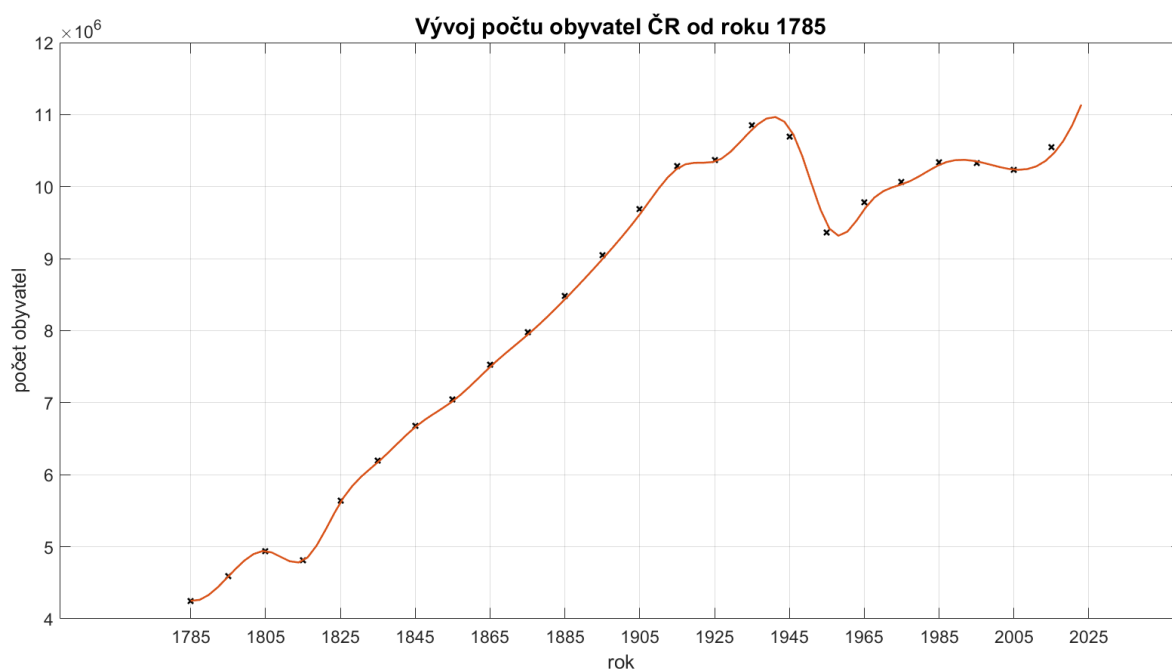
10 Aplikace interpolačních a regresních metod na vývoj počtu obyvatel

Dalším významným a často sledovaným ukazatelem, který vypovídá o (nejen ekonomické) situaci příslušného státu je počet jeho obyvatel. V této kapitole použijeme výše zmíněné metody na vývoj počtu obyvatel na území současné České republiky od roku 1785 do roku 2022 dostupný z ČSÚ, zobrazený v příloze H a podíváme se, zda na těchto datech vykazují stejné vlastnosti (*Obyvatelstvo - roční časové řady, b.r.*).

10.1 Interpolace

Na obrázku 12 můžeme vidět graf vývoje počtu obyvatel měřený po desetiletích interpolovaný kubickým splinem. Porovnáme-li tento spline s hodnotami skutečně naměřenými v jednotlivých letech, dostaneme průměrnou odchylku přibližně 97 677, tedy přibližně 1,17 % z průměrné hodnoty zkoumané veličiny. To znamená, že při zhuštění dat opět dochází k očekávanému zpřesnění informace, resp. použitím méně dat vzrůstá vzdálenost odhadu od skutečných hodnot ve vynechaných bodech. Opět se ale vzhledem k tomu, kolik dat jsme vynechali, jedná o poměrně malou chybu při jejímž akceptování (spokojením se s méně přesným výsledkem) by bylo možné si ušetřit 9 z 10 měření.

Obrázek 12: Kubický spline – vývoj počtu obyvatel ČR



Zdroj: vlastní zpracování

Podíváme-li se na systém z hlediska energie Sobolevových prostorů, dostaneme hodnoty uvedené v tabulce 6. Ty opět potvrzují, že zvolená hustota měření má ne energii velký vliv a při její interpretaci je potřeba brát z úvahu celý kontext, ve kterém je vypočítávána.

Tabulka 6: Energie systému počtu obyvatel ČR z hlediska hustoty měření

energie/měření	po desetiletích	ročně
celková	9 734 554	13 928 923
průměrná	405 606	58 524
maximální	1 326 943	1 169 646
minimální	5 983	483

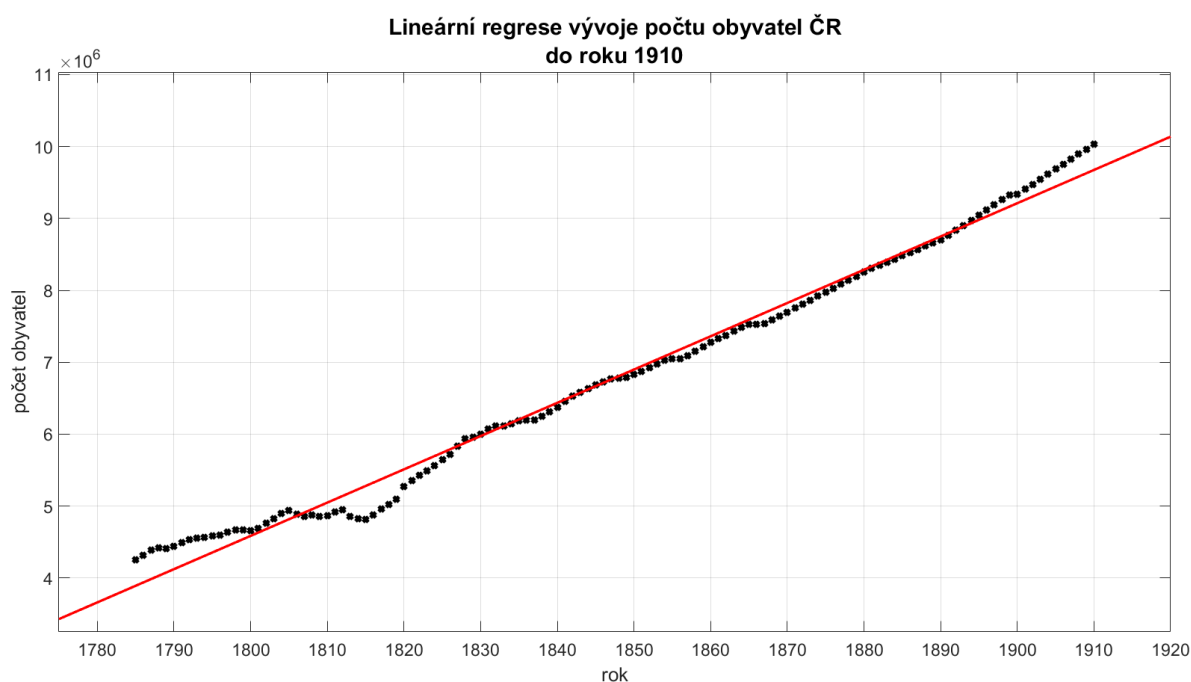
Zdroj: vlastní zpracování

10.2 Lineární regrese

Proložíme-li data do roku 1910 regresní přímkou (obrázek 13), ta bude mít (s užitím optimalizace chyby pomocí E_2) předpis:

$$y(x) = 46\,260x - 78\,681\,949. \quad (39)$$

Obrázek 13: Lineární regrese (E_2) – vývoj počtu obyvatel ČR



Zdroj: vlastní zpracování

Dosažením do rovnice (39) dostaneme odhad hodnoty v roce 1911, který je 9 720 648 obyvatel. Skutečný počet obyvatel v roce 1911 byl 10 099 152. Náš odhad tedy vykazuje odchylku 3,75 %, což opět spadá to pětiprocentní hranice tolerance podle rovnice (36), nicméně vzhledem k hustotě a velkému počtu již dostupných dat není tato hodnota příliš uspokojivá.

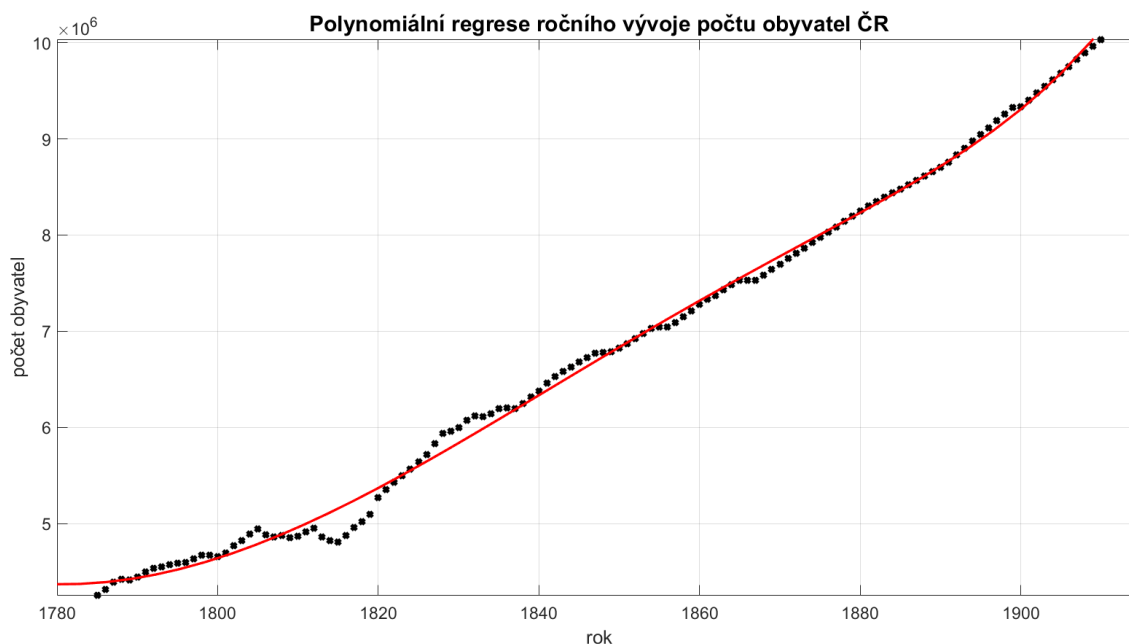
19. století, stejně jako konec 18. a začátek 20., je z hlediska vývoje obyvatelstva mimořádně stabilní a úspěšné období, kdy na území neprobíhaly závažnější politické změny či revoluce, mimořádně krvavé války, ani epidemie, které by měly vývoj obyvatelstva výrazně ovlivnit migrací nebo mortalitou.

10.3 Polynomiální regrese

Jelikož lineární regrese nedokázala ani při tomto počtu dat poskytnout suverénní odhad hledané hodnoty, pojďme se opět podívat, jakého odhadu jsme schopní dosáhnout pomocí polynomiálních funkcí.

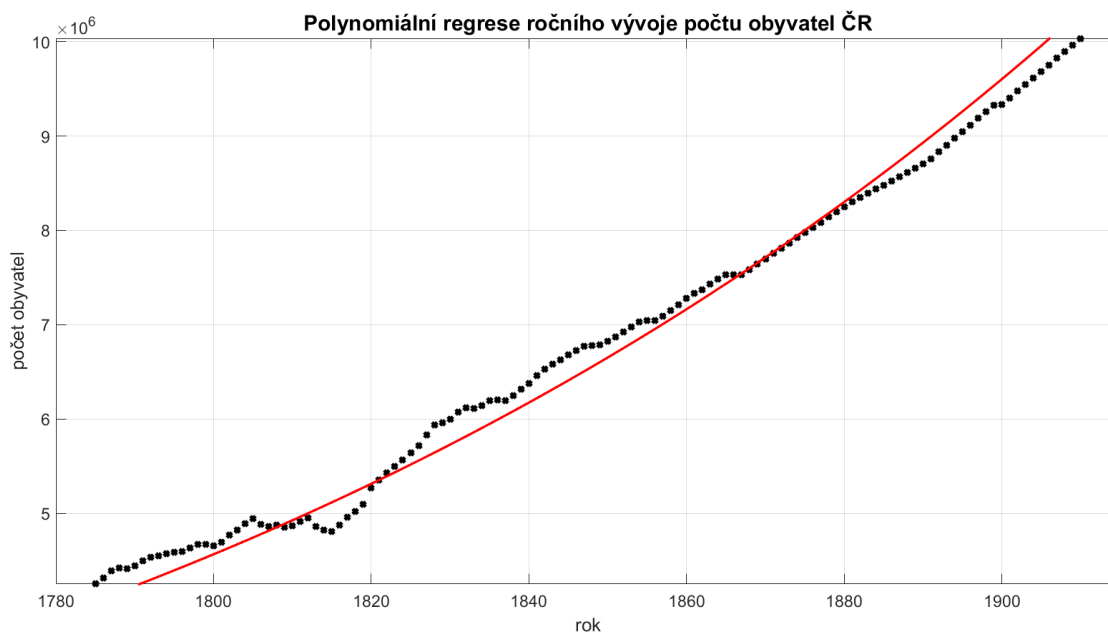
Proložením dat polynomem 21. stupně s optimalizací chyby E_2 (obrázek 14) dostaneme odhadovaný počet obyvatel 10 237 425 s chybou 1,37 %. Polynom 13. stupně optimalizovaný podle E_∞ (obrázek 15) nám dá odhad 10 398 406 obyvatel a odchylku 2,96 %. Opět se tedy ukazuje, že použití častější metody E_2 poskytuje přesnější výsledky.

Obrázek 14: Polynomiální regrese (E_2) – vývoj počtu obyvatel ČR



Zdroj: vlastní zpracování

Obrázek 15: Polynomiální regrese (E_∞) – vývoj počtu obyvatel ČR

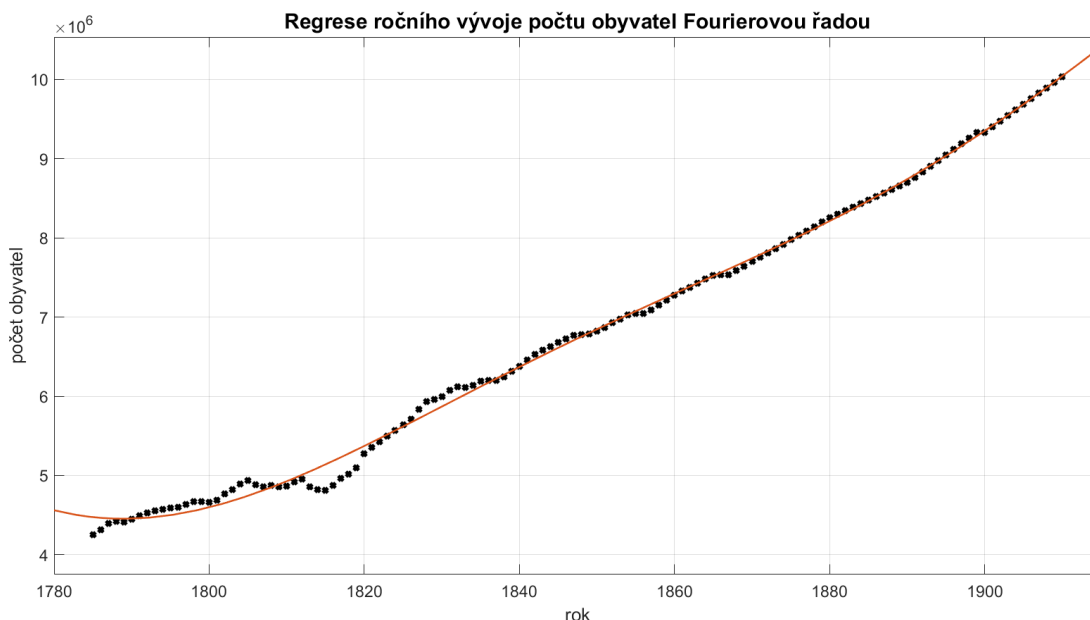


Zdroj: vlastní zpracování

10.4 Regrese Fourierovou řadou

Pro použití Fourierovy řady se i v tomto případě ukázala nejvhodnější (povahou zkoumaným datům nejbližší) řada se dvěma harmonickými složkami s obecným předpisem (38). Po nalezení její optimální polohy (obrázek 16) dostaneme odhad 10 114 172 obyvatel s odchylkou pouhých 0,15 %. Tato metoda se tak pro naši sadu dat ukazuje jako s přehledem nejlepší. Je ale důležité mít na paměti, že pro nalezení a ověření její optimální varianty byla kromě softwarových výpočtů zapotřebí také lidská znalost a základní porozumění funkcím, což tuto metodu dělá méně univerzální.

Obrázek 16: Regrese Fourierovou řadou (E_2) – vývoj počtu obyvatel ČR



Zdroj: vlastní zpracování

10.5 Regrese v čase neobvyklých událostí

Již v předchozí kapitole jsme si ukázali, že i metody, které doposud na dané sadě dat fungovaly velmi dobře, mohou v určitém bodě selhat pod vlivem neočekávané vnější síly, která zapůsobí na systém. Pokusme se nyní z měření proběhlých do roku 1945 předpovídat počet obyvatel v roce 1946 (tabulka 7). Skutečná hodnota naměřená v roce 1946 byla 9 523 266 obyvatel.

Tabulka 7: Různé regrese počtu obyvatel ČR pro rok 1946 a jejich chyby

metoda	odhad počtu obyvatel pro rok 1946	chyba [%]
lin. reg. E_2	11 352 694	19,21
lin. reg. E_∞	11 236 926	17,99
polynom. reg. E_2	11 115 126	16,72
polynom. reg. E_∞	10 835 344	13,78
Fourier. reg.	10 984 423	15,34

Zdroj: vlastní zpracování

Opět se ukazuje, že i navzdory velkému množství dat, která doposud vykazovala jednotný (v tomto případě rostoucí) trend, může nastat situace, kdy se všechny naše metody ukáží být velmi chybové. Všimněme si, že chybovost všech metod se změnila podobným způsobem. To naznačuje, že změna v jejich schopnosti přesného odhadu je skutečně způsobena extrémní neočekávanou změnou v systému, spíše než selháním metod samotných.

V tomto konkrétním případě se jedná o důsledek druhé světové války, během níž se samozřejmě snižoval počet živých a po níž došlo k markantnímu odsunu obyvatel do zahraničí.

Závěr

Tématem této práce je analýza finančních dat metodami prostorů funkcí. Jejím hlavním cílem bylo seznámit čtenáře s analytickými metodami interpolace, extrapolace a regrese a pokusit se na vybraných ekonomických datech zjistit, zda je možné aplikací těchto metod učinit užitečné (akceptovatelně přesné) odhady ve chvílích, kdy nemáme ke skutečným hodnotám přístup.

Po provedení experimentu se ukázalo, že v případě interpolace poskytuje velice vhodnou reprezentaci kubický interpolační spline. Jelikož se jedná o hladkou funkci, nechává uživateli prostor pro další analytické operace, jako jsou derivace a integrace. Navíc je reprezentovatelný konečným množstvím parametrů, což umožňuje jeho velmi přesné uchování a zpracování pomocí výpočetních softwarů. Také se potvrdilo, že podle očekávání se s vyšší hustotou dat zvyšuje přesnost reprezentace dat interpolační funkcí. Jelikož dnes veškeré výpočty mohou probíhat s pomocí počítačů a zpracování velkého množství dat tak není problém, platí pro účely naší analýzy jednoduše „čím více dat, tím lépe.“ Pokud ovšem ještě nemáme data k dispozici a budeme k jejich získání muset provádět měření, bude získání n -krát většího počtu dat zpravidla n -krát náročnější. Jak jsme si ale v této práci ukázali, přínos nových dat v takovém případě může růst pomaleji. Je tedy důležité vhodně určit interval prováděného měření. V obou našich případech se zvolení ročního intervalu ukázalo jako dostatečné.

Z vybraných regresí aplikovaných na „klidných“ datech (měřených v období, kdy nedocházelo k žádným neočekávaným událostem) se ukázala být nejpřesnější regrese Fourierovou řadou. Mírnou nevýhodou této metody je, že vyžaduje složitější kód na nalezení vhodného počtu harmonických složek a určitou míru vzhledu uživatele. Stále se ale jedná o proveditelnou regresi, která poskytuje velmi dobré výsledky. O něco jednodušší je pak použití polynomiální regrese, optimalizované podle E_2 . Tato metoda stále poskytuje dostatečně přesné výsledky a její provedení vyžaduje pouze volbu stupně použitého polynomu, což lze navíc opět provést softwarovým testováním. Nejjednodušší funkcí, kterou lze na regresi dat použít, je funkce lineární. I ta na námi zvolených datech poskytuje s optimalizací E_2 solidní odhady. Její největší výhodou je jednoduchá interpretace a rychlá aplikovatelnost. Je však potřeba mít na paměti, že lineární funkce nám umožňuje zachytit pouze jeden směr trendu dat, pročež nemusí být pro některé datové sady vhodná.

Použití těchto regresních metod na „divoká“ data (měřená v době zásadních, těžko očekávatelných událostí) nám ukázalo, že náhlé změny systému nejsme žádnou z nich schopni dostatečně přesně predikovat. Nejlepší se ukázala být polynomiální regrese s optimalizací E_∞ , která ale stále nedokáže poskytnout odhad dostatečně přesný, aby bylo možné ho považovat za užitečný.

Je třeba si uvědomit, že tato práce zkoumá pouze vlastnosti naměřených dat. Jelikož v reálném světě (a nejenak v ekonomii) jsou vždy důsledky velkých událostí časově opožděny za událostí samotnou, je často možné přidat do našeho analytického procesu informaci, že s vysokou pravděpodobností bude následující zkoumané období tímto zásahem ovlivněno a podle toho zvolit jiné metody, popřípadě si být vědomi toho, že pro nadcházející období bude náš odhad velmi nepřesný.

Summary

This work explores whether it is possible to approximate discrete data such as development of an exchange rate of the foreign exchange market using a continuous mathematical function. Its aim is to get accurate, yet easily manageable representation of data, which could be used to replace some data when missing or to predict their future behaviour. Its methodology is based on mathematical principles of interpolation, extrapolation, and regression. The thesis evaluates various interpolation and regression functions to determine their precision and simplicity, identifying the most appropriate one. It also shows that appropriately chosen functions are very good to work with data in periods of peace and economic stability. However, when any unexpected external factor (such as a war or a major political change) gets involved, it causes the real data to change drastically, which is something that none of the methods is able to work with.

Key words:

approximation, accurate data representation, discrete and continuous data, exchange rate, extrapolation, interpolation, regression, regression function

Seznam použité literatury

EUR RUB Historical Data. (2024, duben 2). Investing.Com.

<https://www.investing.com/currencies/eur-rub-historical-data>

Feistauer, M., & Kučera, V. (2014). *Základy numerické matematiky* (Vyd. 1).

Matfyzpress.

Krisl, T. (2006). *Fourierovy řady* [Bakalářská práce]. Masarykova univerzita v Brně.

Kurzy devizového trhu – měsíční průměry—Česká národní banka. (b.r.). Získáno 2.

duben 2024, z [https://www.cnb.cz/cs/financni-trhy/devizovy-trh/kurzy-](https://www.cnb.cz/cs/financni-trhy/devizovy-trh/kurzy-devizoveho-trhu/kurzy-devizoveho-trhu/prumerne_mena.html?mena=EUR)

[devizoveho-trhu/kurzy-devizoveho-trhu/prumerne_mena.html?mena=EUR](https://www.cnb.cz/cs/financni-trhy/devizovy-trh/kurzy-devizoveho-trhu/kurzy-devizoveho-trhu/prumerne_mena.html?mena=EUR)

Leoni, G. (2017). *A first course in Sobolev spaces* (Second edition). American

Mathematical Society.

Lukeš, J., & Malý, J. (2002). *Míra a integrál* (2. vyd). Karolinum.

Obyvatelstvo—Roční časové řady. (b.r.). Obyvatelstvo - roční časové řady. Získáno 2.

duben 2024, z https://www.czso.cz/csu/czso/obyvatelstvo_hu

Pikal, D. (2020). *Spline funkce a jejich aplikace* [Bakalářská práce]. Univerzita Hradec

Králové.

Sklenák, Vilém. (2001). *Data, informace, znalosti a Internet* (Vyd. 1). C.H. Beck.

Šťastný, P. (2011). *Základní věty matematické analýzy a jejich aplikace* [Diplomová

práce]. Masarykova univerzita.

Seznam obrázků a tabulek

Obrázek 1: Graf Lagrangeovy interpolace.....	19
Obrázek 2: Lineární regrese - příklad vlivu jednoho extrému na jinak konzistentní sadu dat.....	22
Obrázek 3: Souřadnice bodů na přímce.....	25
Obrázek 4: Metoda nejmenších čtverců.....	26
Obrázek 5: Graf lineární regrese dat metodou nejmenších čtverců.....	27
Obrázek 6: Lineární regrese s použitím E_p	29
Obrázek 7: Kubický spline - vývoj kurzu EUR/CZK	32
Obrázek 8: Lineární regrese (E_2) - vývoj kurzu EUR/CZK.....	35
Obrázek 9: Polynomiální regrese (E_2) - vývoj kurzu EUR/CZK.....	37
Obrázek 10: Regrese Fourierovou řadou (E_2) - vývoj kurzu EUR/CZK	38
Obrázek 11: Vývoj měnového kurzu EUR/RUB	39
Obrázek 12: Kubický spline - vývoj počtu obyvatel ČR	42
Obrázek 13: Lineární regrese (E_2) - vývoj počtu obyvatel ČR	43
Obrázek 14: Polynomiální regrese (E_2) - vývoj počtu obyvatel ČR	44
Obrázek 15: Polynomiální regrese (E_∞) - vývoj počtu obyvatel ČR.....	45
Obrázek 16: Regrese Fourierovou řadou (E_2) - vývoj počtu obyvatel ČR.....	46
Tabulka 1: Kubický spline - MATLAB funkce.....	32
Tabulka 2: Energie systému EUR/CZK z hlediska hustoty měření.....	33
Tabulka 3: Regrese polynomem vyššího stupně - MATLAB funkce.....	36
Tabulka 4: Různé regrese kurzu EUR/RUB pro rok 2005 a jejich chyby	40
Tabulka 5: Různé regrese kurzu EUR/RUB pro rok 2015 a jejich chyby	40
Tabulka 6: Energie systému počtu obyvatel ČR z hlediska hustoty měření.....	43
Tabulka 7: Různé regrese počtu obyvatel ČR pro rok 1946 a jejich chyby.....	46

Přílohy

A Měsíční měnový kurz devizového trhu EUR/CZK

rok	leden	únor	březen	duběn	květen	červen	červenec	srpen	září	říjen	listopad	prosinec
1999	35,638	37,715	37,989	37,997	37,692	37,152	36,521	36,415	36,356	36,587	36,403	36,054
2000	36,025	35,709	35,595	36,31	36,555	36,017	35,619	35,356	35,425	35,275	34,617	34,817
2001	35,139	34,64	34,601	34,55	34,382	33,975	33,855	34,034	34,188	33,562	33,325	32,592
2002	32,078	31,789	31,388	30,356	30,558	30,295	29,749	30,796	30,193	30,653	30,756	31,192
2003	31,49	31,645	31,758	31,625	31,391	31,41	31,877	32,289	32,354	31,985	31,974	32,313
2004	32,723	32,857	32,984	32,514	31,974	31,614	31,521	31,634	31,6	31,484	31,287	30,647
2005	30,31	29,961	29,782	30,13	30,216	30,032	30,191	29,592	29,305	29,677	29,261	28,975
2006	28,721	28,409	28,65	28,508	28,271	28,385	28,445	28,193	28,38	28,29	28,03	27,777
2007	27,841	28,231	28,055	28,01	28,231	28,545	28,33	27,858	27,573	27,336	26,731	26,3
2008	26,051	25,376	25,221	25,067	25,098	24,314	23,529	24,286	24,497	24,787	25,183	26,106
2009	27,169	28,459	27,229	26,76	26,738	26,545	25,787	25,649	25,349	25,836	25,827	26,076
2010	26,136	25,976	25,54	25,313	25,666	25,78	25,305	24,807	24,651	24,526	24,637	25,165
2011	24,449	24,276	24,392	24,291	24,383	24,285	24,341	24,273	24,557	24,848	25,453	25,515
2012	25,532	25,041	24,676	24,799	25,322	25,641	25,434	25,02	24,731	24,938	25,365	25,216
2013	25,563	25,476	25,662	25,84	25,889	25,761	25,949	25,818	25,787	25,658	26,925	27,517
2014	27,484	27,443	27,394	27,45	27,44	27,451	27,457	27,816	27,598	27,58	27,666	27,634
2015	27,895	27,608	27,379	27,443	27,396	27,306	27,091	27,041	27,082	27,103	27,04	27,026
2016	27,026	27,039	27,051	27,031	27,026	27,061	27,038	27,025	27,021	27,021	27,032	27,03
2017	27,02	27,02	27,02	26,824	26,564	26,263	26,074	26,101	26,075	25,765	25,536	25,647
2018	25,452	25,319	25,429	25,364	25,644	25,777	25,842	25,683	25,608	25,818	25,934	25,83
2019	25,649	25,725	25,676	25,677	25,77	25,604	25,551	25,802	25,867	25,699	25,53	25,498
2020	25,215	25,05	26,575	27,263	27,268	26,68	26,505	26,166	26,721	27,202	26,463	26,312
2021	26,141	25,875	26,179	25,924	25,558	25,454	25,641	25,468	25,387	25,485	25,4	25,257
2022	24,469	24,439	25,008	24,437	24,748	24,719	24,577	24,568	24,573	24,532	24,367	24,269
2023	23,958	23,712	23,683	23,438	23,604	23,696	23,902	24,112	24,387	24,586	24,489	24,483
2024	24,716	25,232										

B Čtvrtletní měnový kurz devizového trhu EUR/CZK

rok	leden-březen	duben-červen	červenec-září	říjen-prosinec
1999	37,156	37,607	36,428	36,34
2000	35,771	36,289	35,459	34,911
2001	34,798	34,298	34,024	33,202
2002	31,756	30,405	30,247	30,853
2003	31,628	31,477	32,167	32,089
2004	32,859	32,027	31,587	31,126
2005	30,014	30,126	29,677	29,298
2006	28,6	28,384	28,33	28,045
2007	28,037	28,266	27,923	26,829
2008	25,562	24,826	24,092	25,342
2009	27,599	26,677	25,598	25,915
2010	25,868	25,589	24,913	24,786
2011	24,375	24,321	24,388	25,279
2012	25,083	25,261	25,065	25,167
2013	25,568	25,831	25,852	26,657
2014	27,441	27,447	27,618	27,624
2015	27,624	27,38	27,072	27,057
2016	27,039	27,039	27,028	27,028
2017	27,02	26,532	26,084	25,651
2018	25,402	25,599	25,711	25,862
2019	25,682	25,685	25,738	25,58
2020	25,631	27,054	26,465	26,663
2021	26,07	25,638	25,496	25,377
2022	24,653	24,644	24,573	24,387
2023	23,785	23,588	24,136	24,521

C Roční měnový kurz devizového trhu EUR/CZK

rok	leden
1999	35,638
2000	36,025
2001	35,139
2002	32,078
2003	31,49
2004	32,723
2005	30,31
2006	28,721
2007	27,841
2008	26,051
2009	27,169
2010	26,136
2011	24,449
2012	25,532
2013	25,563
2014	27,484
2015	27,895
2016	27,026
2017	27,02
2018	25,452
2019	25,649
2020	25,215
2021	26,141
2022	24,469
2023	23,958
2024	24,716

D Předpisy kubického interpolačního splinu

Pro interval [1999, 2000]:

$$s(x) = -0.4368x^3 + 0.6739x^2 + 0.1499x + 35.6380$$

Pro interval [2000, 2001]:

$$s(x) = -0.4368x^3 + -0.6365x^2 + 0.1873x + 36.0250$$

Pro interval [2001, 2002]:

$$s(x) = 1.2820x^3 + -1.9469x^2 + -2.3961x + 35.1390$$

Pro interval [2002, 2003]:

$$s(x) = -0.0433x^3 + 1.8991x^2 + -2.4439x + 32.0780$$

Pro interval [2003, 2004]:

$$s(x) = -1.7609x^3 + 1.7693x^2 + 1.2246x + 31.4900$$

Pro interval [2004, 2005]:

$$s(x) = 1.6199x^3 + -3.5134x^2 + -0.5195x + 32.7230$$

Pro interval [2005, 2006]:

$$s(x) = -0.2488x^3 + 1.3464x^2 + -2.6866x + 30.3100$$

Pro interval [2006, 2007]:

$$s(x) = -0.7398x^3 + 0.6000x^2 + -0.7402x + 28.7210$$

Pro interval [2007, 2008]:

$$s(x) = 1.5889x^3 + -1.6194x^2 + -1.7596x + 27.8410$$

Pro interval [2008, 2009]:

$$s(x) = -1.7980x^3 + 3.1474x^2 + -0.2315x + 26.0510$$

Pro interval [2009, 2010]:

$$s(x) = 0.5439x^3 + -2.2464x^2 + 0.6695x + 27.1690$$

Pro interval [2010, 2011]:

$$s(x) = 1.1194x^3 + -0.6148x^2 + -2.1917x + 26.1360$$

Pro interval [2011, 2012]:

$$s(x) = -1.5976x^3 + 2.7435x^2 + -0.0629x + 24.4490$$

Pro interval [2012, 2013]:

$$s(x) = 1.4490x^3 + -2.0493x^2 + 0.6313x + 25.5320$$

Pro interval [2013, 2014]:

$$s(x) = -1.2563x^3 + 2.2976x^2 + 0.8796x + 25.5630$$

Pro interval [2014, 2015]:

$$s(x) = 0.1760x^3 + -1.4711x^2 + 1.7061x + 27.4840$$

Pro interval [2015, 2016]:

$$s(x) = 0.7821x^3 + -0.9430x^2 + -0.7081x + 27.8950$$

Pro interval [2016, 2017]:

$$s(x) = -1.1614x^3 + 1.4032x^2 + -0.2479x + 27.0260$$

Pro interval [2017, 2018]:

$$s(x) = 1.4384x^3 + -2.0809x^2 + -0.9255x + 27.0200$$

Pro interval [2018, 2019]:

$$s(x) = -1.2653x^3 + 2.2344x^2 + -0.7721x + 25.4520$$

Pro interval [2019, 2020]:

$$s(x) = 1.2267x^3 + -1.5615x^2 + -0.0992x + 25.6490$$

Pro interval [2020, 2021]:

$$s(x) = -1.6507x^3 + 2.1187x^2 + 0.4580x + 25.2150$$

Pro interval [2021, 2022]:

$$s(x) = 1.4179x^3 + -2.8333x^2 + -0.2566x + 26.1410$$

Pro interval [2022, 2023]:

$$s(x) = -0.2620x^3 + 1.4204x^2 + -1.6695x + 24.4690$$

Pro interval [2023, 2024]:

$$s(x) = -0.2620x^3 + 0.6345x^2 + 0.3855x + 23.9580$$

E Roční měnový kurz devizového trhu EUR/RUB

rok	cena eura [RUB]	rok	cena eura [RUB]
1999	26,254	2012	39,636
2000	27,777	2013	40,783
2001	26,663	2014	47,442
2002	26,366	2015	77,755
2003	34,29	2016	81,789
2004	35,579	2017	64,911
2005	36,516	2018	69,75
2006	34,183	2019	74,89
2007	34,499	2020	70,865
2008	36,32	2021	91,853
2009	45,72	2022	86,973
2010	42,105	2023	76,662
2011	40,799	2024	97,178

F Kód pro zjištění vhodného počtu harmonických složek Fourierovy řady

```
clc; close all;

%% načtení dat z excelu
T = readtable("eur_to_czk_all_time.xlsx");
A = table2array(T);

Y = A(:,2);
X = 1999:1:1999+length(Y)-1;

% počet testovaných harmonických složek
pocet_harmonickych_slozek = 1:1:8;

chyba = zeros(size(pocet_harmonickych_slozek));

% K-násobná křížová validace
k = 5; % Počet skupin pro křížovou validaci
indices = crossvalind('Kfold', length(X), k);

for i = 1:length(pocet_harmonickych_slozek)
    chyby_k = zeros(1, k); % Pole pro uchování chyb pro každou skupinu

    for j = 1:k
        test_indices = (indices == j);
        train_indices = ~test_indices;
        fourier_fit = fit(X(train_indices).', Y(train_indices), ['fourier',
num2str(pocet_harmonickych_slozek(i))]);
```

```

    chyby_k(j) = rms(fourier_fit(X(test_indices)) - Y(test_indices));
end

% Průměrná chyba pro daný počet složek
chyba(i) = mean(chyby_k);
end

for i=1:1:length(pocet_harmonickych_slozek)
    if min(chyba)==chyba(i)
        display(pocet_harmonickych_slozek(i));
    end
end

% Vykreslení závislosti chyby na počtu harmonických složek
plot(pocet_harmonickych_slozek, chyba, '-x');
grid on
xlabel('Počet harmonických složek');
ylabel('Průměrná RMS chyba');
title('Závislost chyby na počtu harmonických složek');

```

G Kód pro regresi Fourierovou řadou

```

clc; close all;

% načtení dat z excelu
T = readtable("eur_to_czk_all_time-do2022.xlsx");
A = table2array(T);

% roční
Y = A(:,2);
X = 1999:1:1999+length(Y)-1;

% Počet harmonických složek pro Fourierovu aproximaci
pocet_harmonickych_slozek = 2; % optimální počet zjištěn scriptem
harmonicke_slozky (příloha F)

% Fitování dat Fourierovou řadou
fourier_fit = fit(X', Y, ['fourier', num2str(pocet_harmonickych_slozek)]);

% Získání předpisu výsledné Fourierovy řady
predpis = formula(fourier_fit);

a0 = fourier_fit.a0;
a1 = fourier_fit.a1;
b1 = fourier_fit.b1;
a2 = fourier_fit.a2;
b2 = fourier_fit.b2;

```

```

w = fourier_fit.w;

fprintf("Předpis výsledné Fourierovy řady: %d + %d*cos(x*%d) + %d*sin(x*%d) +
%d*cos(2*x*%d) + %d*sin(2*x*%d)", a0, a1, w, b1, w, a2, w, b2, w);

odhad2023 = a0 + a1*cos(2023*w) + b1*sin(2023*w) + a2*cos(2*2023*w) +
b2*sin(2*2023*w);
display(odhad2023)

%% Vykreslení
border = 5;

% body
figure ('Name', 'E_inf');
plot(X, Y, 'kx', 'linewidth',3);
grid on;
title('Polynomiální regrese ročního vývoje měnového kurzu EUR/CZK');
xlabel('rok')
ylabel('cena eura (Kč)')
xlim([min(X)-border max(X)+border])
ylim([min(Y)-border max(Y)+border])
hold on;

% polynom
fplot(@(x) a0 + a1*cos(x*w) + b1*sin(x*w) + a2*cos(2*x*w) + b2*sin(2*x*w),
[min(X)-border, max(X)+border], '-', 'linewidth',1.5);

% osy
xline(0,'linewidth',2);
yline(0,'linewidth',2);
hold off;

% osy
xline(0,'linewidth',2);
yline(0,'linewidth',2);
hold off;

```

H Vývoj počtu obyvatel ČR

rok	počet ob.	rok	počet ob.	rok	počet ob.
1785	4250000	1885	8480362	1985	10336742
1786	4318000	1886	8524953	1986	10340737
1787	4392000	1887	8569544	1987	10348834
1788	4418000	1888	8614135	1988	10356359
1789	4414000	1889	8658727	1989	10362257
1790	4444000	1890	8703318	1990	10362740
1791	4497000	1891	8761173	1991	10308682
1792	4531000	1892	8832295	1992	10317807

1793	4553000	1893	8903416	1993	10330607
1794	4570000	1894	8974537	1994	10336162
1795	4588000	1895	9045659	1995	10330759
1796	4598000	1896	9116780	1996	10315353
1797	4633000	1897	9187901	1997	10303642
1798	4668000	1898	9259022	1998	10294943
1799	4670000	1899	9330143	1999	10282784
1800	4659000	1900	9333853	2000	10272503
1801	4692000	1901	9404689	2001	10224192
1802	4767000	1902	9474876	2002	10200774
1803	4823000	1903	9545134	2003	10201651
1804	4894000	1904	9615027	2004	10206923
1805	4941000	1905	9684512	2005	10234092
1806	4883000	1906	9754475	2006	10266646
1807	4857000	1907	9824544	2007	10322689
1808	4876000	1908	9894520	2008	10429692
1809	4853000	1909	9964789	2009	10491492
1810	4870000	1910	10035575	2010	10517247
1811	4916000	1911	10099152	2011	10496672
1812	4951000	1912	10157344	2012	10509286
1813	4859000	1913	10221343	2013	10510719
1814	4826000	1914	10283486	2014	10524783
1815	4810000	1915	10285882	2015	10542942
1816	4878000	1916	10221815	2016	10565284
1817	4961000	1917	10128304	2017	10589526
1818	5021000	1918	10004335	2018	10626430
1819	5093000	1919	9921710	2019	10669324
1820	5272791	1920	9978420	2020	10700155
1821	5353785	1921	10002030	2021	10500850
1822	5426885	1922	10112730	2022	10759525
1823	5494633	1923	10198370		
1824	5567415	1924	10277770		
1825	5641433	1925	10369760		
1826	5715138	1926	10442610		
1827	5831825	1927	10495940		
1828	5933743	1928	10549221		
1829	5959643	1929	10597761		
1830	5996788	1930	10648057		
1831	6071798	1931	10702208		
1832	6117669	1932	10750003		
1833	6114791	1933	10791313		
1834	6139975	1934	10826082		
1835	6191643	1935	10853125		
1836	6200250	1936	10872519		
1837	6197244	1937	10888540		
1838	6248539	1938	10877442		
1839	6314283	1939	11105990		
1840	6378071	1940	11159539		
1841	6459128	1941	11129373		
1842	6531908	1942	11054018		
1843	6578355	1943	11034846		
1844	6630222	1944	11109341		
1845	6682437	1945	10692912		

1846	6723727	1946	9523266
1847	6768064	1947	8765230
1848	6778356	1948	8893104
1849	6789915	1949	8892613
1850	6826465	1950	8925122
1851	6870370	1951	9023170
1852	6925780	1952	9125183
1853	6976739	1953	9220908
1854	7029810	1954	9290617
1855	7044048	1955	9365969
1856	7045616	1956	9442040
1857	7091652	1957	9513758
1858	7151837	1958	9574650
1859	7214655	1959	9618554
1860	7277801	1960	9659818
1861	7328855	1961	9588016
1862	7373633	1962	9621808
1863	7429933	1963	9668741
1864	7483274	1964	9730019
1865	7528107	1965	9785102
1866	7532419	1966	9826188
1867	7533238	1967	9854241
1868	7584956	1968	9877632
1869	7643553	1969	9896695
1870	7698830	1970	9805157
1871	7754152	1971	9830602
1872	7809472	1972	9868379
1873	7864793	1973	9919519
1874	7920115	1974	9994761
1875	7975435	1975	10062366
1876	8030757	1976	10128220
1877	8086077	1977	10189312
1878	8141398	1978	10245686
1879	8196719	1979	10296489
1880	8252040	1980	10326792
1881	8301997	1981	10303208
1882	8346588	1982	10314321
1883	8391180	1983	10322823
1884	8435770	1984	10330481