

# VYUŽITÍ WAVELETŮ PŘI ANALÝZE ČASOVÝCH ŘAD 2. PRAKTICKÁ ČÁST

## USING WAVELETS BY TIME SERIES ANALYSIS 2. PRACTICAL PART

**Vratislava Mošová**

Moravská vysoká škola Olomouc  
vratislava.mosova@mvso.cz

### **Abstrakt:**

*Waveletová transformace je moderní matematický nástroj, který se využívá nejen při zpracování zvukového signálu či obrazu, ale lze ho také použít při studiu časových řad. V článku je na příkladu prezentován waveletový rozklad v kombinaci s Boxovou - Jenkinsovou metodologií jako efektivní prostředek prognózování časové řady.*

### **Klíčová slova:**

*Časové řady, ARIMA modely, autokorelační funkce, parciální autokorelační funkce, diskrétní waveletová transformace, waveletové koeficienty.*

### **Abstract:**

*Wavelet transformation is a recent mathematical tool that is used not only in audio signal processing or image processing but it can be also used in the study of time series. In this paper, the wavelet decomposition combined with Box-Jenkins methodology is presented as an effective means of predicting in time series.*

### **Keywords:**

*Time series, ARIMA models, autocorrelation function, partial correlation function, discrete wavelet transform, wavelet coefficients.*

**JEL:** C53

## **1 Úvod**

Tento článek je druhou částí textu „Využití waveletů při analýze časových řad“ (viz [4]). Zatímco cílem publikované první části bylo seznámit čtenáře s teoretickou základnou problematiky, je předložena druhá část ukázkou použití prezentované teorie při zkoumání ekonomických časových řad.

Následující text se skládá ze dvou částí. V první z nich je objasněno propojení Boxovy-Jenkinsovy teorie s waveletovou transformací a jsou uvedeny poznámky k realizaci obou přístupů. Ve druhém odstavci je prezentována ukázka postupu při prognózování časové řady nejprve pomocí čisté Boxovy-Jenkinsovy metodologie a pak pomocí metody modifikované prostřednictvím waveletů.

## 2 Několik připomínek k realizaci použitých metod

Boxova-Jenkinsova metodologie byla vytvořena pro řady s navzájem závislými pozorováními. Reziduální složky takových řad mohou být tvořeny korelovanými náhodnými veličinami. Tato závislost se zkoumá prostřednictvím nástrojů pro korelační analýzu.

Aby model konstruovaný na základě Boxovy-Jenkinsovy metodologie odpovídal realitě, je třeba, aby zpracovávaná časová řada měla alespoň 50 členů. V grafech ACF a PACF pak kontrolujeme 1/7 hodnot z celkového rozsahu řady.

Identifikace vhodného modelu ARIMA začíná vyhodnocením spojnicového grafu. Vytipuje se, zda je řada stacionární a jestli obsahuje sezonní nebo cyklickou složku. Takto získané postřehy jsou ovšem značně subjektivní.

Přesnější závěry vyplynou z grafů ACF a PACF. Pokud je první hodnota ACF hodnota blízká jedné, není řada stacionární. Odstranění nestability v rozptylu lze dosáhnout pomocí Boxovy-Coxovy transformace

$$F(x) = \begin{cases} x^\lambda & \text{pro } \lambda > 0 \\ \ln x & \text{pro } \lambda = 0 \end{cases}$$

Nestacionarita ve střední hodnotě se potlačí pomocí diferencování. Řád difference je možné odhadnout ze spojnicového grafu diferencí, popř. z grafů ACF a PACF pro příslušné difference.

Přítomnost cyklů nejlépe odhalíme pomocí periodogramu. Pokud se v periodogramu na pozicích různých od nuly nacházejí výrazné vrcholy, vyskytují se v časové řadě cykly s periodami, které přísluší nejvyšším hodnotám periodogramu.

Z uvedených pozorování se pak odvíjí volba modelu. Zvolený model je třeba opět konfrontovat s grafy ACF a PACF. V případě že hodnoty v těchto grafech vyběhají z vyznačených mezí spolehlivosti, jsou nesystematické složky modelu korelované. Jedná se o nevhodný model a je třeba vybrat jiný.

Jestliže je k dispozici víc než jeden ARIMA model, preferuje se ten, který má méně parametrů a jehož hodnoty parametrů jsou nižší. Po nalezení nejvhodnějšího modelu, je možné provést požadovanou prognózu.

Při predikci časové řady s využitím waveletů se nejprve najde vhodný wavelet, který je co nejvíc podobný zadané časové řadě a který odpovídá sledovaným cílům. Tak například pro hladká data se používají hladké škálové funkce. Pro data s jemnými detaily se vybírají škálové funkce s krátkým nosičem. Pro rekonstrukci dat tvořených polynomem  $N$ -tého stupně se používají wavelety, které mají  $N-1$  nulových momentů.

V dalším kroku se s využitím vhodně zvoleného waveletu provede rozklad časové řady na aproximace a detaily. Pomocí Boxovy-Jenkinsovy metodologie se vyhledá vhodný ARIMA model jak pro aproximace tak pro detaily rozložené řady. Zvlášť pro aproximace a zvlášť pro detaily se provede příslušná predikce. Celková prognóza řady se obdrží jako součet predikovaného vektoru pro aproximace s predikovaným vektorem pro detaily.

### 3 Modifikace modelů ARIMA prostřednictvím waveletů - příklad

#### Zadání

V Tabulce 1 jsou uvedeny měsíční hodnoty míry inflace ČR pro léta 2000-2014.

Tabulka 1: Hodnoty zadané řady

	1	2	3	4	5	6	7	8	9	10	11	12
2000	2.1	2.2	2.3	2.4	2.5	2.6	2.9	3.1	3.3	3.6	3.8	3.9
2001	4.	4.	4.	4.1	4.2	4.3	4.5	4.6	4.7	4.7	4.7	4.7
2002	4.6	4.6	4.6	4.5	4.3	3.9	3.5	3.1	2.7	2.4	2.1	1.8
2003	1.5	1.1	0.8	0.5	0.3	0.2	0.2	0.1	0.	0.	0.1	0.1
2004	0.3	0.5	0.8	1.	1.2	1.4	1.7	2.	2.2	2.5	2.7	2.8
2005	2.8	2.7	2.6	2.6	2.5	2.4	2.2	2.1	2.	2.	1.9	1.9
2006	2.	2.1	2.2	2.3	2.4	2.5	2.6	2.7	2.8	2.7	2.6	2.5
2007	2.4	2.3	2.2	2.2	2.1	2.1	2.1	2.	2.	2.2	2.5	2.8
2008	3.4	3.9	4.3	4.7	5.	5.4	5.8	6.1	6.4	6.6	6.5	6.3
2009	5.9	5.4	5.	4.6	4.1	3.7	3.1	2.6	2.1	1.6	1.3	1.
2010	0.9	0.8	0.7	0.6	0.6	0.6	0.8	0.9	1.1	1.2	1.4	1.5
2011	1.6	1.7	1.7	1.8	1.8	1.9	1.9	1.9	1.8	1.9	1.9	1.9
2012	2.1	2.2	2.4	2.6	2.7	2.8	2.9	3.1	3.2	3.3	3.3	3.3
2013	3.2	3.	2.8	2.7	2.5	2.3	2.2	2.	1.8	1.6	1.5	1.4
2014	1.3	1.1	1.									

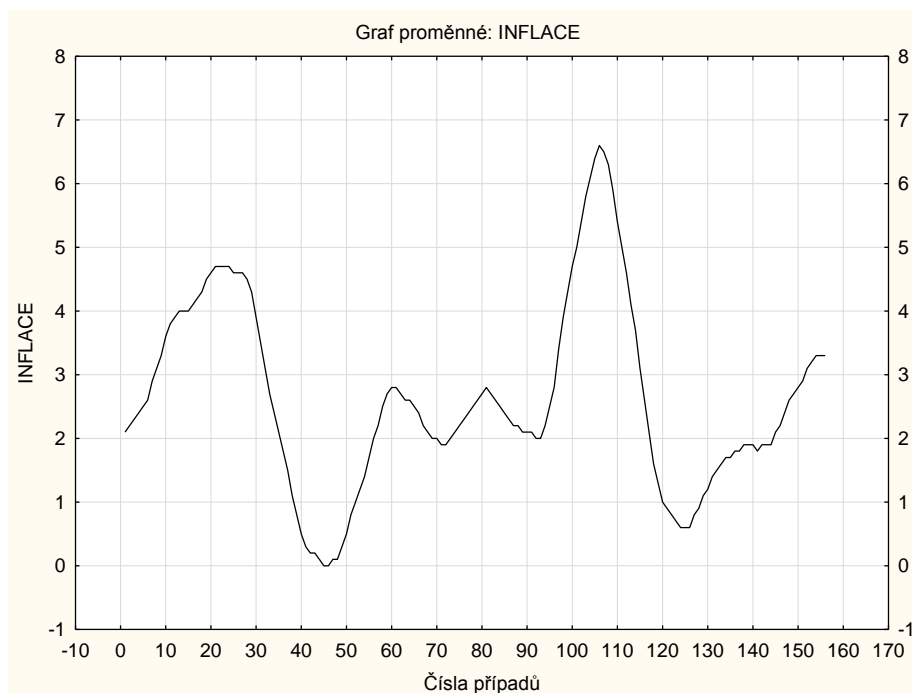
Pro řadu hodnot od ledna 2000 do prosince 2012 vyhledejte vhodný model ARIMA. Proveďte prognózu pro zbytek řady. Výsledek konfrontujte s výsledkem, který získáte pomocí kombinace waveletového rozkladu kombinovaného s ARIMA modely.

#### Řešení:

A) Predikce pomocí ARIMA modelu:

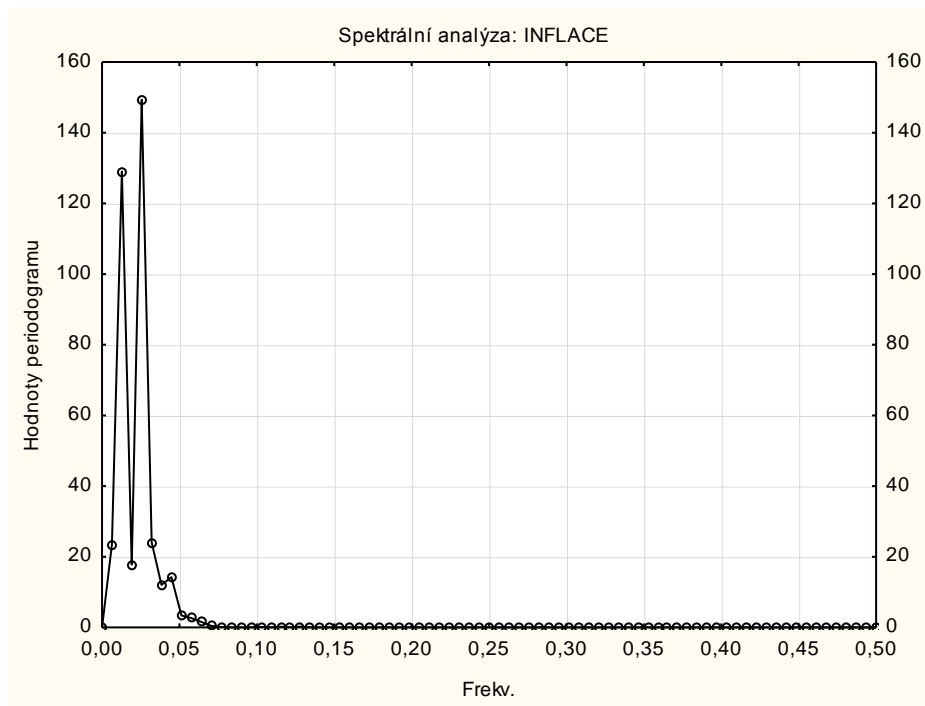
Úlohu budeme řešit v prostředí programů Statistica 10 a Wolfram Mathematica 9. Ze spojnicového grafu (viz Obrázek 1) je patrné, že řada je nestacionární.

Obrázek 1: Spojnicový graf řady Inflation



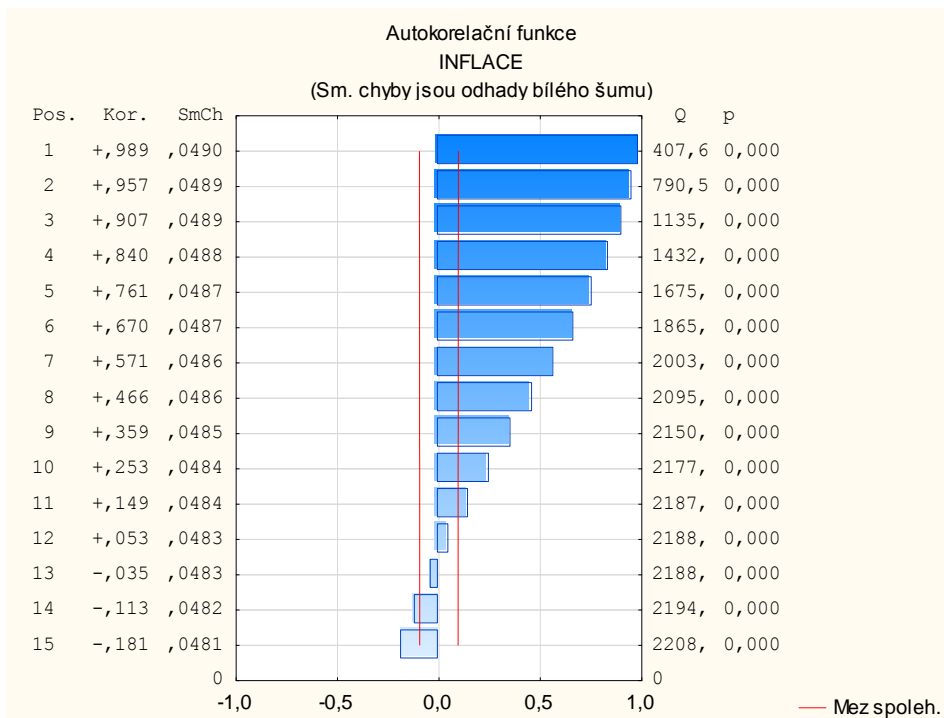
Z periodogramu (viz Obrázek 2) je vidět, že v řadě jsou 3 cykly.

Obrázek 2: Periodogram řady Inflace

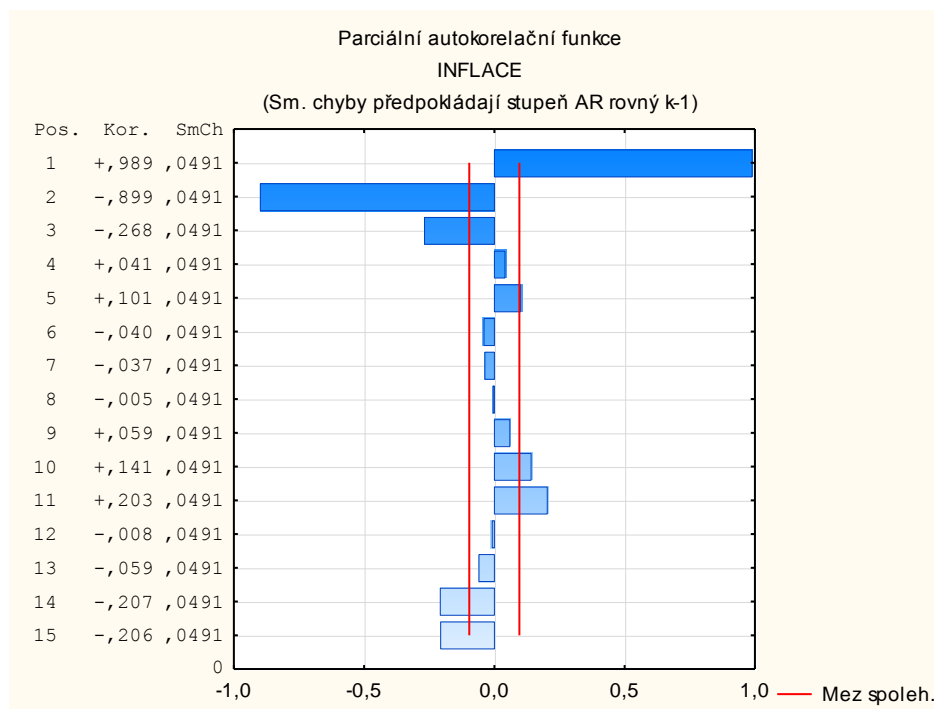


Grafy ACF (viz Obrázek 3) a PACF (viz Obrázek 4) naznačují, že řada je nestacionární a že se pravděpodobně jedná o autoregresní model AR(3).

Obrázek 3: ACF pro řadu Inflace



Obrázek 4: PACF pro řadu Inflace



Z výše uvedeného plyne, že v řadě je třeba provést transformaci, diferencovat a hledat sezónní model. Koeficient pro Box-Coxovu transformaci spočtený programem Statistica je  $\lambda \approx 0,5$ . Na základě grafů ACF a PACF pro 2. diferenci původní řady usoudíme, že tato diference stačí ke stacionarizaci původní řady.

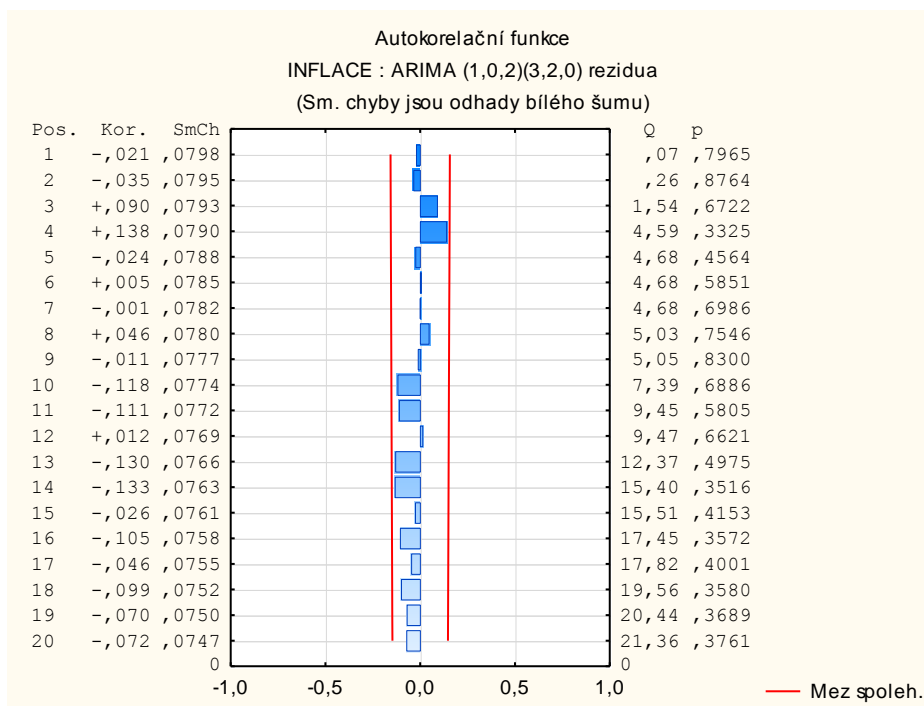
Zadanou řadu zkusíme odhadnout modelem ARIMA(1,0,2)(3,2,0). Parametry modelu spolu s dalšími údaji jsou uvedeny v tabulce (viz Tabulka 2):

Tabulka 2: Parametry modelu ARIMA(1,0,2)(3,2,0)

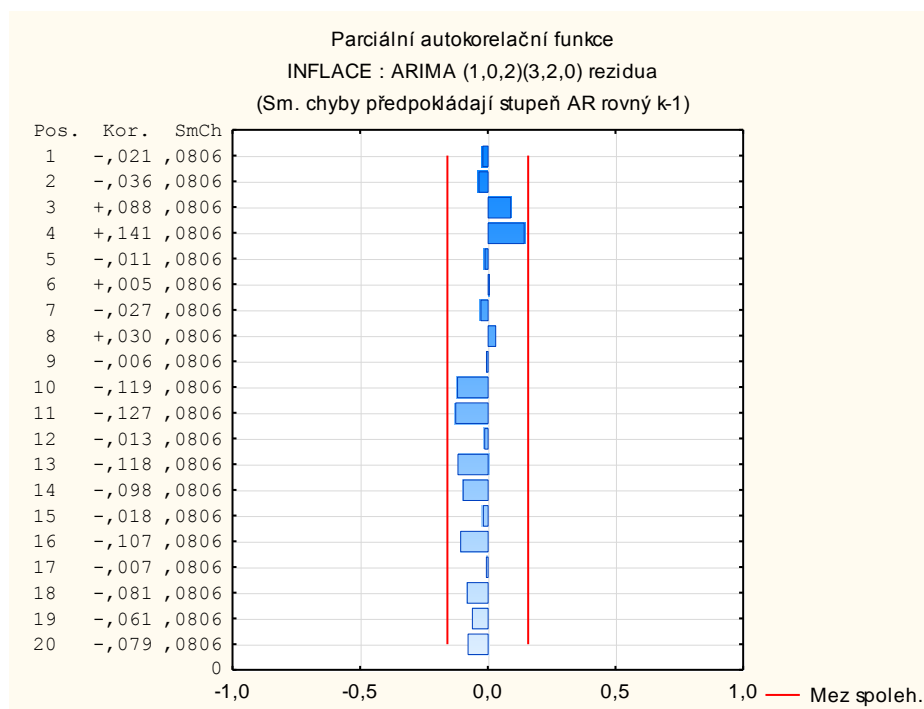
Vstup: INFLACE (ARIMAtab.sta) Transformace: $x^{,5000,2}D(1)$ Model:(1,0,2)(3,2,0) Sezónní posun: 12 PČ Rezid. = ,00329						
	Param.	Asympt. - SmCh	Asympt. - t(148)	p	Dolní - 95% spol	Horní - 95% spol
(1)	-0,463346	0,401741	-1,15334	0,250628	-1,25724	0,330543
(1)	-0,054280	0,388594	-0,13968	0,889100	-0,82219	0,713630
(2)	0,276491	0,171677	1,61053	0,109412	-0,06276	0,615746
s(1)	-0,127412	0,087855	-1,45024	0,149107	-0,30102	0,046201
s(2)	0,021660	0,084570	0,25612	0,798212	-0,14546	0,188782
s(3)	-0,067797	0,088421	-0,76675	0,444452	-0,24253	0,106934

Na následujících dvou obrázcích Obrázku 5 a Obrázku 6 se hodnoty ACF a PACF pohybují uvnitř mezi spolehlivosti. To znamená, že model je vhodně zvolený,

Obrázek 5: ACF pro model ARIMA(1,0,2)(3,2,0)

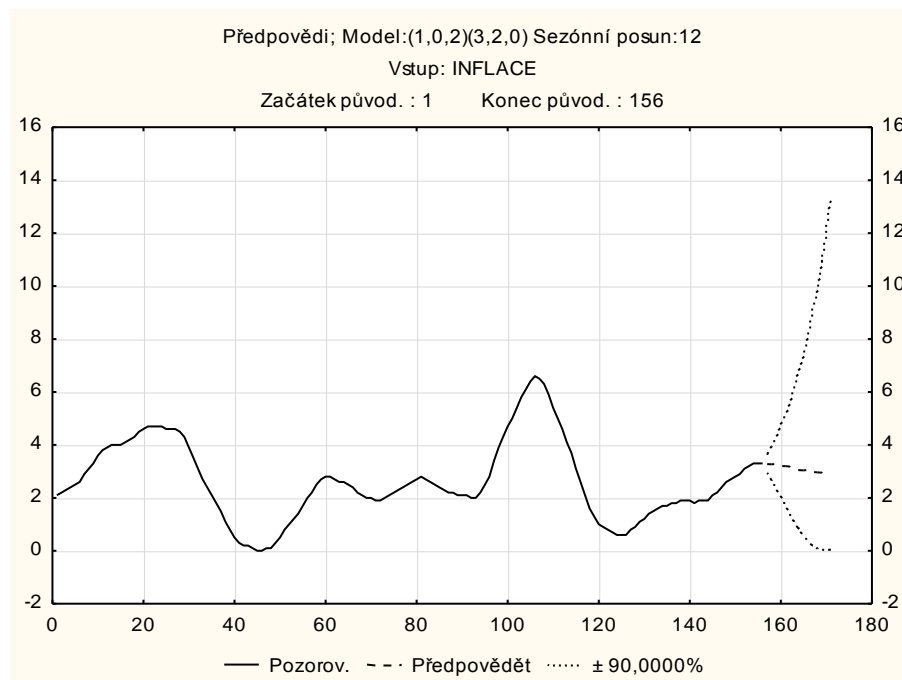


Obrázek 6: PACF pro model ARIMA(1,0,2)(3,2,0)



Graf na Obrázku 7 a v Tabulce 3 je zachycena předpověď, která vychází z modelu ARIMA(1,0,2)(3,2,0). Dále je možné z hodnot v Tabulce 1 a Tabulce 3 určit velikost odmocniny ze střední kvadratické chyby: RMSE=1,20801.

Obrázek 7: Graf prognózy Inflace - model ARIMA(1,0,2)(3,2,0)



Tabulka 3: Hodnoty prognózy Inflace - model ARIMA(1,0,2)(3,2,0).

Pořadí	157	158	159	160	161	162	163	164
Prognóza	3,27630	3,26771	3,24345	3,22398	3,20165	3,18239	3,13182	3,08467

Pořadí	165	166	167	168	169	170	171
Prognóza	3,03740	3,00876	2,98008	2,96232	2,95306	2,93955	2,93938

### B) Predikce pomocí waveletů:

V programu Mathematica se provede waveletový rozklad pomocí Daubechiové waveletu Db3 a obrzčí se dva soubory (viz Tabulka 4 a Tabulka 5).

Tabulka 4: Hodnoty řady Aproximace

2.61051, 2.07221, 2.16841, 2.4481, 2.50979, 2.63189, 2.84345, 3.09829, 3.33792, 3.60084, 3.77563, 3.92223, 3.98143, 3.98091,  
 4.03986, 4.08686, 4.19308, 4.32518, 4.46515, 4.62686, 4.68246, 4.70394, 4.70458, 4.66913, 4.64309, 4.60193, 4.56961, 4.53496,  
 4.2668, 3.89912, 3.52296, 3.07183, 2.72778, 2.39326, 2.09196, 1.82061, 1.47351, 1.11174, 0.796709, 0.486669, 0.323159, 0.229285,  
 0.137974, 0.0959022, 0.0445145, 0.0125085, 0.0494309, 0.123778, 0.294672, 0.531483, 0.753923, 1.01136, 1.21319, 1.40625, 1.68155,  
 1.98186, 2.23963, 2.49973, 2.67669, 2.81485, 2.79288, 2.67579, 2.64425, 2.58759, 2.49271, 2.37886, 2.23937, 2.07602, 2.01188,  
 1.97607, 1.9291, 1.90117, 1.97932, 2.11034, 2.19716, 2.30117, 2.39766, 2.49432, 2.61008, 2.73326, 2.7421, 2.7078, 2.62633, 2.49247,  
 2.39663, 2.28742, 2.22095, 2.17042, 2.1342, 2.11607, 2.05252, 1.97965, 2.05865, 2.19309, 2.47195, 2.85434, 3.33526, 3.9219, 4.307,  
 4.66632, 5.0466, 5.40239, 5.76668, 6.12533, 6.38314, 6.59629, 6.51744, 6.28273, 5.91751, 5.3928, 5.00266, 4.58122, 4.12844, 3.6619,  
 3.14011, 2.58504, 2.09568, 1.61284, 1.28168, 1.02508, 0.871925, 0.813792, 0.693933, 0.603291, 0.595998, 0.631325, 0.751475,  
 0.936515, 1.07383, 1.23022, 1.37151, 1.51026, 1.60458, 1.6772, 1.73356, 1.76814, 1.83121, 1.89447, 1.88984, 1.85942, 1.86773,  
 1.87323, 1.89675, 1.93071, 2.05429, 2.2228, 2.39533, 2.60012, 2.70356, 2.77942, 2.9305, 3.08993, 3.2178, 3.34411, 3.22682, 3.0018

Tabulka 5: Hodnoty řady Detaily

-0.510514, 0.12779, 0.131587, -0.048096, -0.00978644, -0.0318903, 0.0565534, 0.00170805, -0.0379235, -0.00083962, 0.0243681, -0.0222281, 0.0185669, 0.019091, -0.0398561, 0.0131357, 0.00692325, -0.0251831, 0.0348531, -0.0268575, 0.0175406, -0.00393531, -0.0045795, 0.0308679, -0.0430891, -0.00192871, 0.0303901, -0.0349649, 0.0331985, 0.000877169, -0.0229597, 0.0281749, -0.0277758, 0.00674014, 0.008037, -0.0206075, 0.0264874, -0.0117393, 0.00329113, 0.0133311, -0.0231587, -0.0292848, 0.062026, 0.00409784, -0.0445145, -0.0125085, 0.0505691, -0.0237782, 0.00532784, -0.031483, 0.046077, -0.011361, -0.0131941, -0.00624616, 0.0184468, 0.0181434, -0.0396316, 0.000265644, 0.0233128, -0.0148476, 0.00712228, 0.0242105, -0.0442525, 0.0124126, 0.00728865, 0.0211398, -0.0393659, 0.0239775, -0.0118849, 0.0239323, -0.0291017, -0.00116803, 0.0206826, -0.0103432, 0.00284238, -0.00117188, 0.00234375, 0.00568476, -0.0100812, -0.0332615, 0.0578953, -0.00779552, -0.0263298, 0.00752988, 0.00337007, 0.0125789, -0.0209482, 0.0295795, -0.0342004, -0.0160655, 0.047477, 0.0203503, -0.0586475, 0.00691476, 0.0280461, -0.0543391, 0.0647445, -0.0218997, -0.00699834, 0.0336812, -0.0466047, -0.00238979, 0.0333196, -0.0253287, 0.0168637, 0.00370697, -0.0174372, 0.0172711, -0.0175079, 0.00720147, -0.00265928, 0.0187791, -0.0284367, 0.0381028, -0.0401058, 0.0149602, 0.00431745, -0.0128407, 0.018318, -0.0250751, 0.0280754, -0.0137922, 0.0060669, -0.00329113, 0.00400192, -0.0313251, 0.0485249, -0.0365151, 0.0261719, -0.0302237, 0.0284866, -0.0102643, -0.0045795, 0.0228018, -0.0335647, 0.0318566, -0.0312125, 0.00552687, 0.0101601, 0.0405754, -0.0677275, 0.0267662, 0.00325358, -0.03071, 0.0457116, -0.0228018, 0.00466693, -0.000120336, -0.00356166, 0.0205784, -0.0305017, 0.0100688, -0.017798, -0.0441126, 0.0731831, 0.298204}

Při určování modelů vhodných k prognózování řad Aproximace a Detaily se postupuje obdobně jako při prognózování souboru Inflace.

A) *Prognóza pro Aproximace*: Na základě spojnicového grafu, periodogramu, ACF a PACF lze soudit, že v řadě je třeba provést transformaci, diferencovat a hledat sezonní model. Koeficient pro Box-Coxovu transformaci je  $\lambda \cong 0,7$ . Zadanou řadu zkusíme odhadnout modelem ARIMA(1,0,0)(0,1,0).

Hodnoty předpovědi pro ARIMA (1,0,0)(0,1,0) jsou uvedeny v Tabulce 6.

Tabulka 6: Hodnoty prognózy Aproximace - model ARIMA(1,0,0)(0,1,0)

Pořadí	157	158	159	160	161	162	163	164
Prognóza	2,80411	2,63022	2,47708	2,34205	2,22285	2,11750	2,02430	1,94175

Pořadí	165	166	167	168	169	170	171
Prognóza	1,86858	1,80364	1,74597	1,69471	1,64911	1,60851	1,57234

B) *Prognóza pro detaily*: Ze spojnicového grafu, ACF a PACF příslušných k souboru detailů se zjistí, že se jedná o stacionární soubor. Když se rozhodneme pro model ARIMA(2,0,2), obdržíme hodnoty předpovědi pro detaily uvedené v Tabulce 7.

Tabulka 7: Hodnoty prognózy Detaily - model ARIMA(2,0,2)

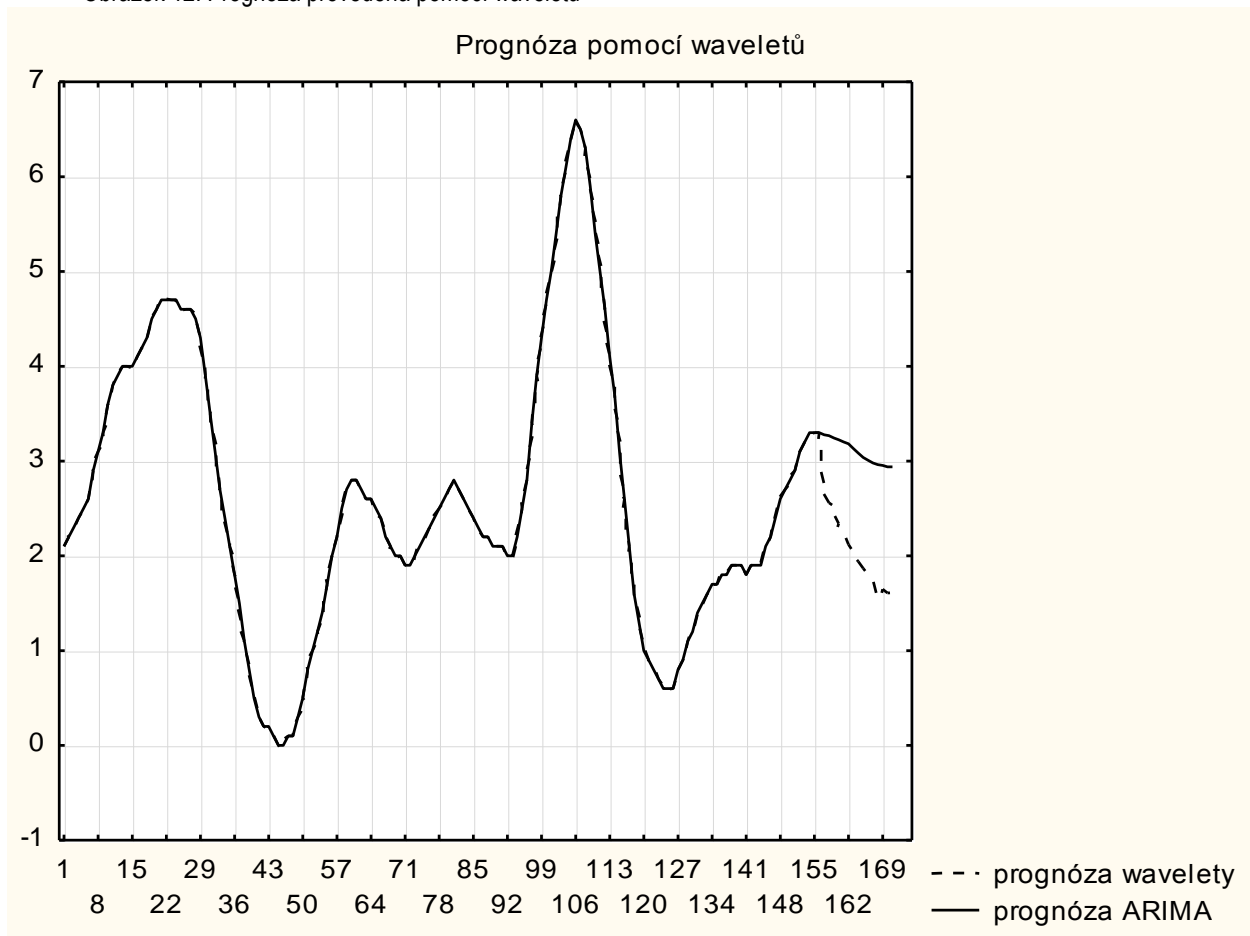
Pořadí	157	158	159	160	161	162	163	164
Prognóza	-0,1258	-0,0712	0,0366	-0,0023	-0,0039	0,0014	0,000041	-0,0002

Pořadí	165	166	167	168	169	170	171
Prognóza	0,00005	0,000007	-0,000008	0,000002	0,000001	0	0

C) *Výslednou prognózu* obdržíme sečtením vektorů hodnot prognózovaných pro aproximace a hodnot prognózovaných pro detaily. Graficky je situace zachycena na Obrázku 12 a v Tabulce 10 jsou uvedeny příslušné prognózované hodnoty. Z hodnot v Tabulce 1 a v Tabulce 10 se určí velikost střední kvadratické chyby: RMSE=0,330203.



Obrázek 12: Prognóza provedená pomocí waveletů



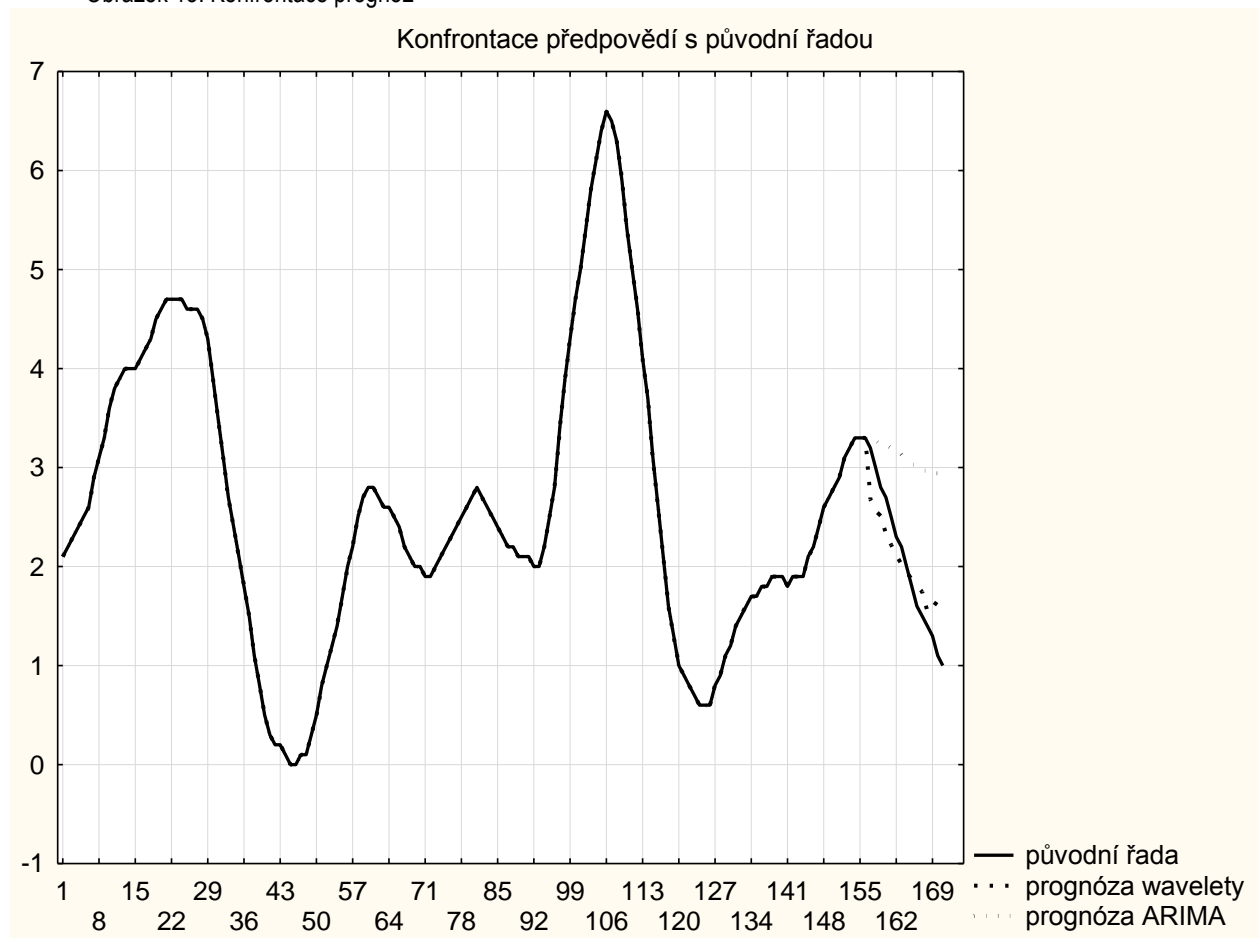
Tabulka 10: Hodnoty prognózy provedené pomocí waveletů

Pořadí	157	158	159	160	161	162	163	164
Prognóza	2,67830	2,55906	2,51369	2,33978	2,21906	2,11891	2,02434	1,94157

Pořadí	165	166	167	168	169	170	171
Prognóza	1,86863	1,80365	1,74596	1,5548	1,64911	1,60851	1,57234

Na Obrázku 13 je vybraný model  $ARIMA(1,0,2)(3,2,0)$  konfrontován s předpovědí sestavenou s pomocí waveletu Db3.

Obrázek 13: Konfrontace prognóz



Obrázek 13: Konfrontace prognóz

Připomeňme ještě, že chyba odhadu  $RMSE=1,20801$  při použití  $ARIMA(1,0,2)(3,2,0)$  je vyšší než  $RMSE=0,330203$  při použití metody s waveletem. Z  $RMSE$  je vidět, že předpověď s využitím waveletu může poskytnout lepší prognózu než klasický model  $ARIMA$ .

#### 4 Závěr

Wavelety jsou moderní matematický aparát, který se používá při řešení praktických problémů jako je zpracování signálu i při řešení teoretických otázek např. v numerické matematice. Vzhledem ke schopnosti waveletů elegantně odstranit korelovanost v datech, lze je využít při zpracování a predikci časové řady. V tomto článku wavelety vystupují spolu s Boxovými-Jenkisovými modely jako nástroj k prognózování časových řad a na příkladu je demonstrováno, že tímto způsobem je možné dosáhnout zlepšení v přesnosti prognózy.

#### Poděkování

Tento příspěvek vznikl s finanční podporou a v rámci řešení projektu GAČR P403/12/1811: Vývoj nekonvenčních modelů manažerského rozhodování v podnikové ekonomice a veřejné ekonomii.

## Literatura

- [1] Artl, J., Artlová, M, Rublíková, E. *Analýza ekonomických časových řad s příklady*. VŠE Praha 2004. ISBN 978-80-247-3243-5.
- [2] Jansen, M., Oonix P. *Second Generation Wavelets and Applications*. Springer – Verlag London 2005. ISBN 1-85233-916-0.
- [3] Najzar, K. *Základy teorie waveletů*, Karolinum Praha, 2004. ISBN 80-246-0957-6.
- [4] Mošová, V. *Využití waveletů při analýze časových řad - 1. Teoretická část*. EMI 4/2014, MVŠO, 2014. ISSN 1804-1299 (Print), ISSN 1805-353X (Online).
- [5] Siegel, A. F. *Practical Business Statistics*, 2011. ISBN 9780123852083.
- [6] Švec, M. *Waveletové transformace*, UJEP Ústí nad Labem, 2008. ISBN 80-246-0957-6.