

ČESKÁ ZEMĚDĚLSKÁ UNIVERZITA V PRAZE PROVOZNĚ

EKONOMICKÁ FAKULTA

Katedra ekonomiky



Prognostické metody

Seminární práce

Autor: Miloš Uldrich

Cvičící: Ing. Lukáš Čechura, Ph.D.

ČT 12:15 (su)

© 2009 ČZU v Praze

Obsah

1. Cíl	- 1 -
2. Základní charakteristiky časové řady	- 1 -
3. Analýza sezonní složky	- 3 -
4. Určení trendové složky	- 5 -
5. Určení řádu integrace.....	- 7 -
6. Určení řádu AR(p) a MA(q)	- 9 -
7. Prognóza cen pohonných hmot	- 10 -
7.1. ARIMA (p, d, q).....	- 10 -
7.2. ADL model	- 11 -
7.3. VAR model	- 16 -
8. Závěr	- 18 -
9. Použité zdroje.....	- 19 -
10. Přílohy.....	- 19 -

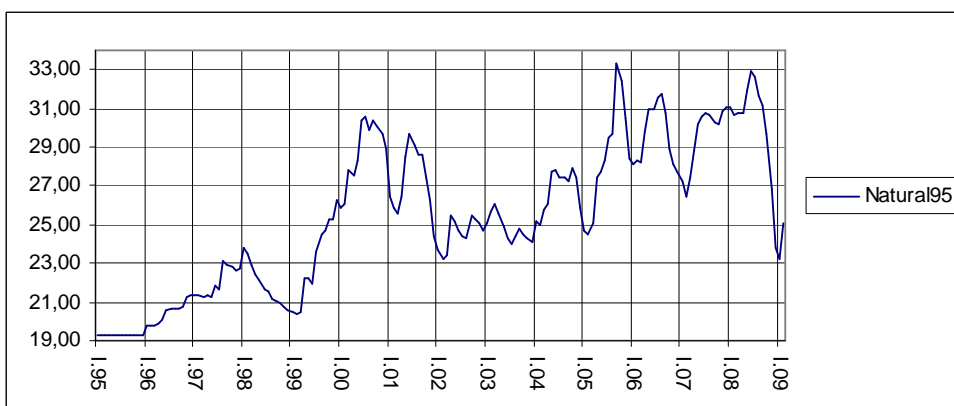
1. Cíl

Cílem práce je porovnat různé přístupy k modelování časových řad. Provést předpovědi hodnot pomocí jednorozměrné a vícerozměrné časové řady a výsledky konfrontovat pomocí vybraného kritéria.

2. Základní charakteristiky časové řady

Bude provedena analýza časové řady, která zachycuje vývoj cen zvolených pohonných hmot od ledna 1995 do února 2009 po měsíčních intervalech. Podkladem jsou průměrná měsíční sledování cen pohonných látek pro Českou republiku, která zveřejňuje Český statistický úřad. Průměrné měsíční ceny zvolených paliv za celou Českou republiku od ledna 1995 do února 2009 ukazuje následující **graf č. 1**.

Graf č. 1: Průběh ceny benzínu Natural 95 v Kč/l v ČR.



Zdroj: ČSÚ

Největší zaznamenaný růst ceny benzínu nastal v září 2005, cena tehdy stoupla o 3,69 Kč/l. Trh s ropou tehdy ovlivnila přírodní katastrofa, hurikán Katrina. Ten vyřadil kromě jiného z provozu řadu amerických rafinerií a zdejší firmy zareagovaly zvýšenou poptávkou po pohonných hmotách. Aby předešly nedostatku, začaly skupovat paliva na trzích v Amsterdamu a Rotterdamu, které jsou cenotvorné i pro prodejce pohonných hmot v České republice. Tempo růstu je pro toto období charakterizováno nejprudším zrychlením. Průměrná cena benzínu „vystřelila“ na 33,33 Kč/l.

Největší meziměsíční pokles cen sledovaného paliva nastal mezi listopadem a prosincem roku 2008. Cena Naturalu 95 se propadla o 2,95 Kč/l na 26,07 Kč/l. Jde zároveň o nejvýznamnější zpomalení ve sledované časové řadě. Příčinou byla krize na newyorské burze, která vyvolala následující celosvětovou ekonomickou krizi. Cena ropy těsně před krachem na burze (1. 7. 2008)

dosáhla na 147 USD/barel. Ekonomická krize vyvolala nedostatek poptávky po pohonných hmotách a nabídka pohonných hmot se ocitla v převaze. OPEC musel reagovat výrazným snížením těžby ropy. Cena za barel se v prosinci 2008 propadla na 40 USD/barel a tento pokles také následovala spotřebitelská cena benzínu.

Vypočtené základní charakteristiky uvádí **tabulka č. 1**. Průměrná cena benzínu za zvolené období je 25,47 Kč/l. Maximální ceny dosáhl benzin v roce 2005 po zmiňovaném hurikánu, a to 33,33 Kč/l. Směrodatná odchylka ukazuje na relativně vysokou variabilitu zvolené časové řady.

Tabulka č. 1: Základní charakteristiky zvolených časových řad

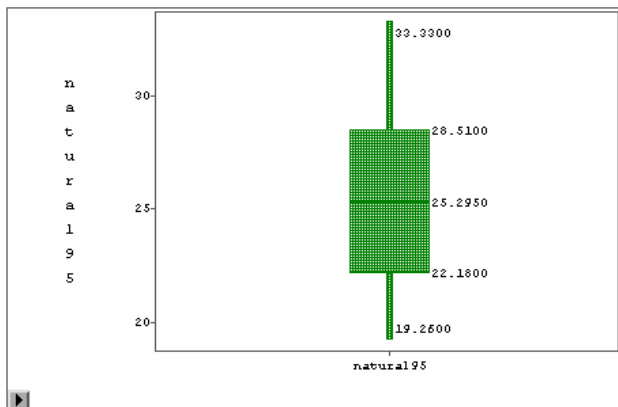
základní charakteristiky	Natural 95
<i>Průměr (Kč/l)</i>	25,47
<i>Maximum (Kč/l)</i>	33,33
<i>Minimum (Kč/l)</i>	19,26
<i>Rozptyl</i>	14,73
<i>Směrodatná odchylka (Kč/l)</i>	3,84
<i>Variační koef. (%)</i>	15,07

Zdroj: vlastní výpočty

Jako další ukazatel pro základní analýzu slouží krabičkový graf – box plot.

Graf č. 2 (box plot) ukazuje kvartilové rozdělení Naturalu 95. Z **grafu č. 2** je vidět, že časová řada neobsahuje žádná extrémní pozorování. Odlehlými pozorováními jsou maximální hodnoty cen obou komodit, tyto hodnoty vždy reagovaly na výjimečnou situaci na trhu (hurikán, krize), takové hodnoty mohou zkreslovat odhady parametrů regresních funkcí, a proto je někdy vhodné aproximovat je náhradní hodnotou. Z grafu lze také vyčíst, že průběh časové řady u benzínu má přibližně normální rozdělení, což lze potvrdit histogramem v **příloze č. 1**.

Graf č. 2: Krabičkový graf zvolené časové řady



Zdroj: vlastní výpočty

Základní charakteristiky, vizuální analýza a box plot, ukázaly, že vybraná statistika je v určitých obdobích zkreslena tzv. „šoky“ v ekonomice. Průměr a rozptyl upozornily na větší variabilitu zvolené řady. Obě informace zhoršují podmínky pro modelování časových řad.

3. Analýza sezonní složky

Je obecně známo, že cena pohonných hmot v letní sezoně stoupá a naopak na přelomu roku bývá nejnižší. Zvolená časová řada tedy pravděpodobně obsahuje významnou sezonní složku. Pro ověření vlivu jednotlivých měsíců existuje více postupů. Jako první byla v programu SAS zkoumána časová řada cen benzínu testem analýzy rozptylu, kterým se dá sezonnost relativně přesně identifikovat.

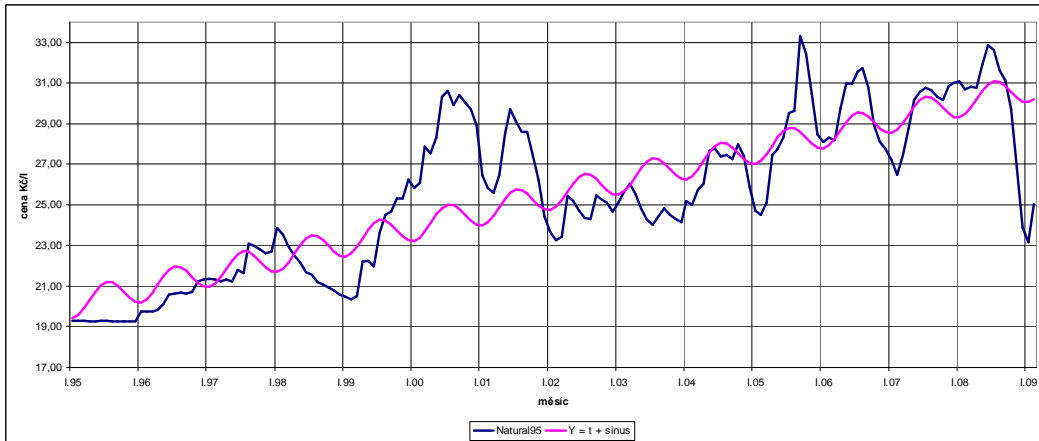
The ANOVA Procedure

Závislá proměnná: Natural 95

Zdroj	DF	Součet čtverců	Průměrný kvadrát	F hodnota	Pr > F
Model	25	2222.609454	88.904378	45.41	<.0001
Chyba	144	281.943405	1.957940		
Korigovaný součet	169	2504.552859			
Zdroj	DF	Type I SS	Průměrný kvadrát	F hodnota	Pr > F
mesice	11	95.432641	8.675695	4.43	<.0001
rok	14	2127.176814	151.941201	77.60	<.0001

V řádku „mesice“ je P-hodnota (Pr > F) 0,001 a je menší než zvolená hladina významnosti, a proto lze říci, že model zobrazuje statisticky významný vliv jednotlivých měsíců, respektive jednotlivé měsíce se od sebe statisticky významně odlišují. Výsledek analýzy rozptylu zároveň ukazuje na významné meziroční změny u cen benzínu. Sezonnost si lze také představit jako průběh funkce sinus. Další závislost budeme modelovat funkcí sinus s periodou dvanáct, jehož grafický průběh, společně se skutečnými cenami benzínu Natural 95, ukazuje **graf č. 3**.

Graf č. 3: Modelování sezonnosti na funkci sinus u benzínu Natural 95



Zdroj: vlastní výpočty

Vzorec funkce sinus je následující

$$\sin\left(\frac{2\pi * t}{12}\right). \quad (3.1)$$

Graf č. 3 zachycuje jednak průběh cen benzínu Natural 95, ale také průběh funkce $Y_t = t + \sin$, kde cena benzínu Y_t je vysvětlována trendem modelovaným na průběhu funkce sinus. Metodou nejmenších čtverců program SAS odhadl parametry u konstanty, časového vektoru a proměnné sinus v této podobě:

$$Y_t = 20,022 + 0,0638t - 0,711\sin. \quad (3.2)$$

Celý výstup z programu SAS si lze prohlédnout v následující tabulce:

Tabulka č. 2: Panel výsledků u zvolené funkce pro benzin Natural 95

Model Equation							
natural95 = 20.0220 + 0.0638 t - 0.7110 sinus_d							
Summary of Fit							
Mean of Response	25.5440	R-Square	0.6691				
Root MSE	2.2066	Adj R-Sq	0.6651				
Analysis of Variance							
Source	DF	Sum of Squares	Mean Square	F Stat	Pr > F		
Model	2	1624.2418	812.1209	166.80	<.0001		
Error	165	803.3773	4.8690				
C Total	167	2427.6191					
Type III Tests							
Source	DF	Sum of Squares	Mean Square	F Stat	Pr > F		
t	1	1605.4644	1605.4644	329.73	<.0001		
sinus_d	1	42.3436	42.3436	8.70	0.0037		
Parameter Estimates							
Variable	DF	Estimate	Std Error	t Stat	Pr > t	Tolerance	Var Inflation
Intercept	1	20.0220	0.3485	57.45	<.0001	.	0
t	1	0.0638	0.0035	18.16	<.0001	0.9970	1.0030
sinus_d	1	-0.7110	0.2411	-2.95	0.0037	0.9970	1.0030

Zdroj: vlastní výpočty

Z výstupu je vidět, že je model jako celek statisticky významný, a pro identifikaci sezonnosti je také důležité, že parametr u proměnné sinus vyšel statisticky významný, což ukazuje řádek „sinus_d“ v tabulce „Type III“, kde je P-hodnota 0,0037 menší než hladina významnosti $\alpha = 0,05$.

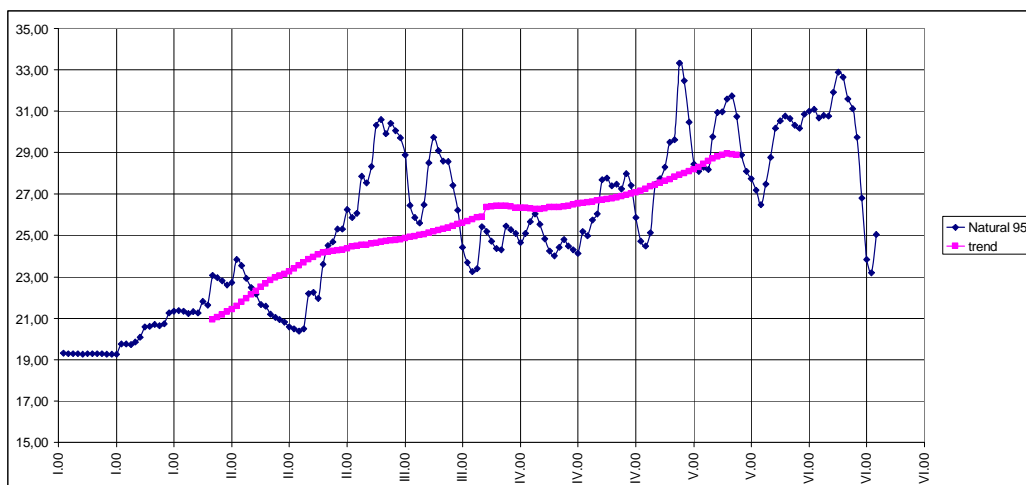
Významnost poukazuje na to, že zvolená funkce (3.2) průběh sezonního kolísání vystihuje. Průběh skutečných cen je však hodně zkreslen dalšími vlivy na trhu. Přibližně od roku 2001 lze sezonní propady vysledovat i z vizuální analýzy **grafu č. 3**, ze které je vidět, že cena dosahuje své minimální hodnoty vždy na přelomu roku, což je v souladu s průběhem vzorce (3.2).

Pro cenu benzínu budeme dále sezonní složku považovat za významnou a v odhadech budou použity sezonně očištěné časové řady.

4. Určení trendové složky

Pokud chceme modelovat trendovou složku, je třeba potvrdit nebo vyvrátit vliv sezonních vlivů, o které by se musel trend (ve formě sezonních indexů) případně korigovat. U benzínu se ukázal být tento vliv jako významný. **Graf č. 4** ukazuje vyhlazený střednědobý trend u benzínu Natural 95 určený pomocí klouzavých průměrů.

Graf č. 4: Střednědobý trend u benzínu Natural 95



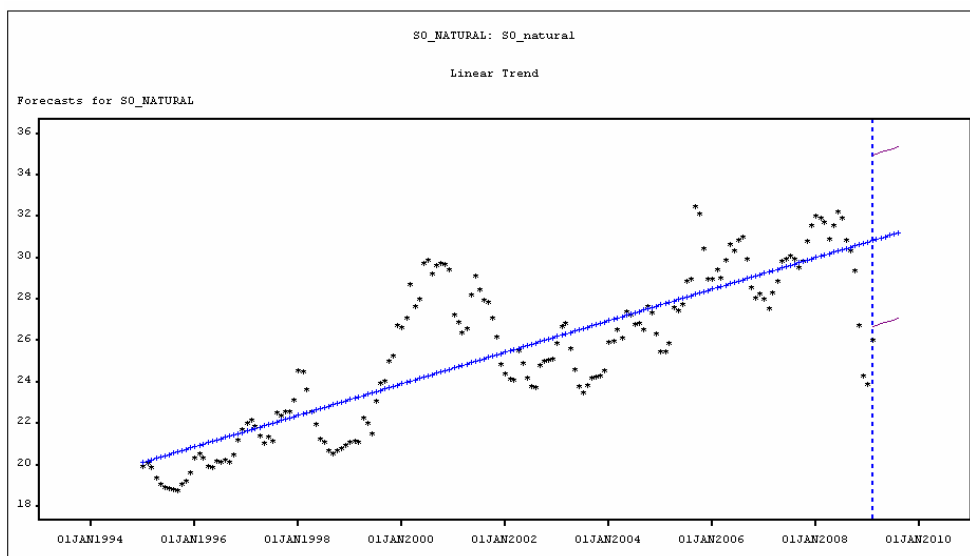
Zdroj: vlastní výpočty

Cena komodity v čase, i přes významné propady, stále roste. Nebýt současné hospodářské krize, započaté v létě 2008, ceny by stále stouply. Obecně se předpokládá, že až skončí hospodářská recese, cena paliv se vrátí na svou předcházející úroveň. V budoucnu bude ropa zcela jistě opět překonávat hranice 60, 80 a 100 USD/barel. Ceny zvolených pohonných hmot porostou také. Otázkou zůstává, jak rychle se celosvětové hospodářství vzpamatuje a ceny pohonných hmot

se vrátí na svou původní úroveň, tj. před hospodářskou krizí. Různé prognózy se od sebe významně odlišují. Nelze říci, zda za několik měsíců nebo několik let.

Sezonně očištěná časová řada byla aproximována přímkou, která ukazuje dlouhodobou tendenci v jejím vývoji. Tento průběh zachycuje následující **graf č. 5**.

Graf č. 5: Lineární aproximace u cen benzínu Natural 95



Zdroj: vlastní výpočty

Tvar zvolené přímky má následující podobu:

$$Y_t = 20,05 + 0,063t. \quad (4.1)$$

Oba parametry jsou statisticky významné od nuly, koeficient determinace R^2 vyšel 0,68 a střední čtvercová chyba $RMSE = 2,1$. Tyto výsledky neukazují na nijak dobrou predikční schopnost modelu, což je dáno především jeho charakterem. Funkce (4.1) byla použita k určení intervalové a bodové prognózy s délkou 6 měsíců. Budoucí hodnoty cen by se měly v příštích letech pohybovat v intervalu, který zobrazuje **příloha č. 2**.

Reziduální složka modelu v tomto případě však netvoří bílý šum, odhad hodnot parametrů může být proto zkreslen a významnost parametrů může být pouze zdánlivá. Aproximace funkcí (4.1) pro zvolené časové řady není proto, vzhledem k charakteru dat, až tak vhodná. **Graf č. 5** pouze ukazuje to, že dlouhodobý vývoj cen je rostoucí, a potvrzuje předchozí úsudek o přibližně rostoucím trendu.

5. Určení řádu integrace

Z výše uvedených důvodů vyplývá, že zvolená časová řada má pravděpodobně nestacionární průběh. Časová řada byla rozdělena na segmenty (jednotlivé roky) a v nich byl spočítán průměr, který se již na první pohled významně liší. U stacionárních řad se jednotlivé průměry od sebe neliší. Pro ověření byla použita analýza rozptylu a autokorelační funkce. Výsledky analýzy rozptylu u benzínu jsou následující:

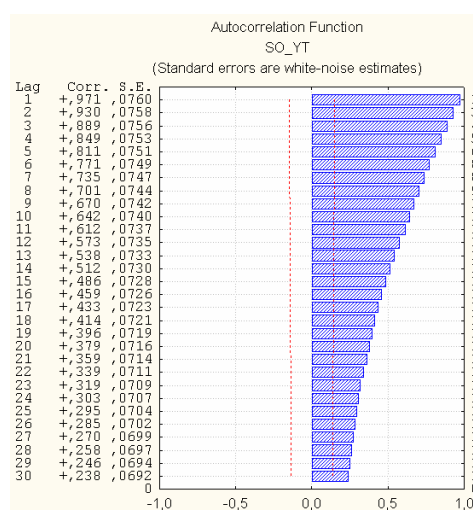
The ANOVA Procedure

Závislá proměnná: Natural 95

Zdroj	DF	Součet čtverců	Průměrný kvadrát	F hodnota	Pr > F
Model	14	2130.791034	152.199360	63.12	<.0001
Chyba	155	373.761825	2.411367		
Korigovaný součet	169	2504.552859			
Zdroj	DF	Anova SS	kvadrát	hodnota	Pr > F
rok	14	2130.791034	152.199360	63.12	<.0001

Z výstupu je vidět, že P-hodnota (Pr > F) v řádku „Model“ je mnohem menší než zvolená hladina významnosti $\alpha = 0,05$. Model je proto jako celek významný. Stejný význam má P-hodnota v řádku „rok“, a jednotlivé roky se proto od sebe statisticky významně liší. Časová řada tak není stacionární. Tuto informaci ještě podpoříme korelogramem v následujícím grafu:

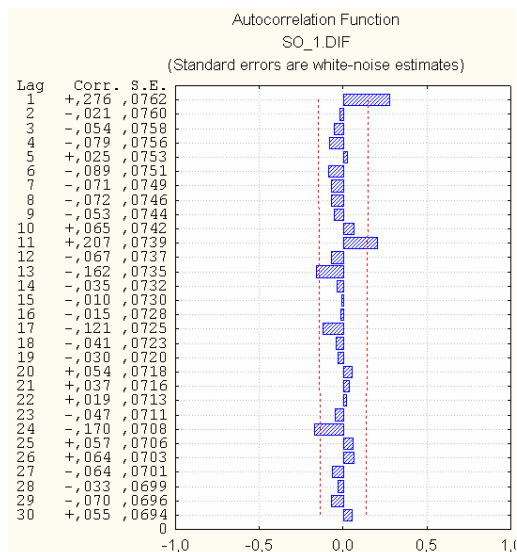
Graf č. 6: Autokorelační funkce pro benzin Natural 95



Zdroj: vlastní výpočty

Z grafu je vidět, že korelační koeficient zde nekonverguje k nule (což u stacionární časové řady očekáváme) a jeho hodnoty neustále přesahují konfidenční pás. Nestacionarita může být způsobena pravděpodobně rostoucím trendem cen a „šoky“ na trhu pohonných hmot. U zvolené časové řady provedeme první diferenci a předchozí postup zopakujeme. Z následujícího **grafu č. 7** je vidět, že korelační koeficient diferencované sezonně očištěné ceny má již relativně optimální průběh.

Graf č. 7: Autokorelační funkce prvních diferencí benzínu Natural 95



Zdroj: vlastní výpočty

Hodnota korelačního koeficientu již téměř nepřesahuje konfidenční pás. Občasné přesahy mohou způsobovat nadměrné rozdíly mezi obdobími, které jsou způsobeny výše uvedenými „šoky“ na trhu pohonných hmot. Z diferencované řady byly již tyto výše zmíněné „šoky“ na trhu odstraněny, aby nezpůsobovaly nepřiměřený rozdíl v daném období, a proto byla pro ověření v programu Statistika opět spočítána analýza rozptylu.

V programu Statistika byl spočten test analýzy rozptylu, který má určit, zda zkoumaná časová řada je po první diferenci již stacionární. Výsledky zobrazuje **tabulka č. 3**.

Tabulka č. 3: Panel výsledků analýzy rozptylu pro zvolené časové řady

Efekt	SČ	Stupně volnosti	PČ	F	p
Abs. člen	0,24901	1	0,249010	0,423221	0,516304
rok	6,77444	14	0,483889	0,822423	0,643972
Chyba	90,60890	154	0,588370		

Zdroj: vlastní výpočty

Výstupy ukazují, že u benzínu se jednotlivé roky statisticky významně neliší, protože P-hodnota (0,64) je větší než zvolená hladina významnosti $\alpha = 0,05$. Z předchozích výpočtů tedy vyplývá, že zvolená časová řada je integrována řádu jedna I(1), tj. že pro stacionaritu, kterou některé přístupy vyžadují, musíme provést, respektive nastavit první diferenci. Např. u Boxovy-Jenkinsovy metodologie by se již nepracovalo z modelem ARMA (p, q), ale z modelem ARIMA (p, d, q), kde $d = 1$.

6. Určení řádu AR (p) a MA (q)

Proces klouzavých součtů MA (q)

Proces klouzavých součtů MA (q), někdy chybně nazýván proces klouzavých průměrů, je jedna z možností modelování dynamiky stacionárních časových řad. Proces generuje minulé informace náhodné složky a na náhodnou složku reaguje postupně. Náhodná informace v časové řadě je využita do hodnoty řádu q . Obecně lze MA ($q = n$) lze zapsat jako:

$$Y_t = e_t + \theta_1 e_{t-1} + \theta_2 e_{t-2} + \dots + \theta_n e_{t-n}, \quad (6.1)$$

kde θ jsou odhadnuté parametry u zpožděných reziduí, jejich počet závisí na volbě q . Je-li zvolen řád $q = 1$, lze říci, že model má paměť pouze jedno období. Proces MA je vždy stacionární s nulovou střední hodnotou, rozptylem a autokorelační funkcí (autokorelační funkce má tedy bod useknutí k rovný řádu modelu q).

Z grafu č. 7 volíme řád $q = 1$, tedy MA (1).

Autoregresní proces AR (p)

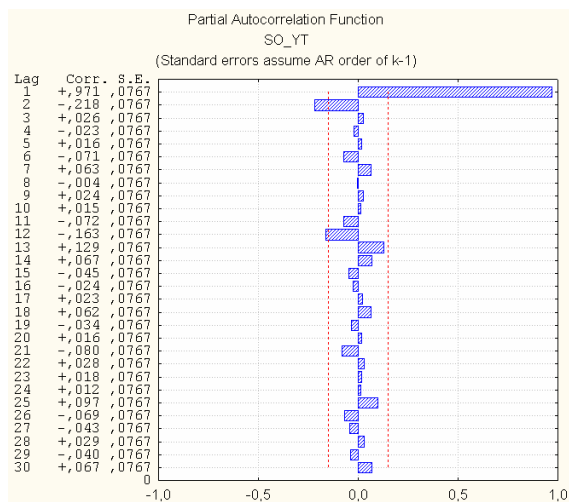
Autoregresní proces AR (p) je další možný přístup modelování dynamiky stacionárních časových řad. Vyjadřuje Y_t jako funkci několika předcházejících pozorování. Náhodná informace je zde oproti MA (q) absorbována v plné výši a hodnota vysvětlující proměnné reaguje na svou minulou hodnotu $B_1 Y_{t-1}$ ve výši parametru B_1 . U náhodné složky požadujeme vlastnosti bílého šumu. Lze říci, že model AR(p) má nekonečnou paměť, ale se zvyšující se hodnotou korelačního koeficientu K se snižuje vliv minulého období. Autoregresní proces řádu p má tvar:

$$Y_t = \beta_1 y_{t-1} + \beta_2 y_{t-2} + \dots + \beta_p y_{t-p} + e_t, \text{ tj. } Y_t - \beta_1 y_{t-1} - \beta_2 y_{t-2} - \dots - \beta_p y_{t-p} = e_t. \quad (6.2)$$

Hodnost p určujeme z parciální autokorelační funkce a je rovna hodnotě k , tedy bodu useknutí korelačního koeficientu.

Z grafu č. 8 parciální autokorelační funkce určíme řád $p = 2$, tedy AR (2).

Graf č. 8: Autokorelační funkce prvních diferencí benzínu Natural 95



Zdroj: vlastní výpočty

Na základě těchto poznatků budeme dále konstruovat smíšený model ARIMA (2, 1, 1).

7. Prognóza cen pohonných hmot

Cílem práce je však vzájemně porovnat schopnosti předpovědních metod, a proto se omezíme pouze na bodovou ex post prognózu, kde budeme ověřovat předpovědní schopnosti na již známých datech. Jako porovnávací kritérium je zvolen součet čtvercových chyb SSE.

$$SSE = \sum_{t=n+1}^{n+h} (Y_t - \hat{y}_t)^2 \quad (7.1)$$

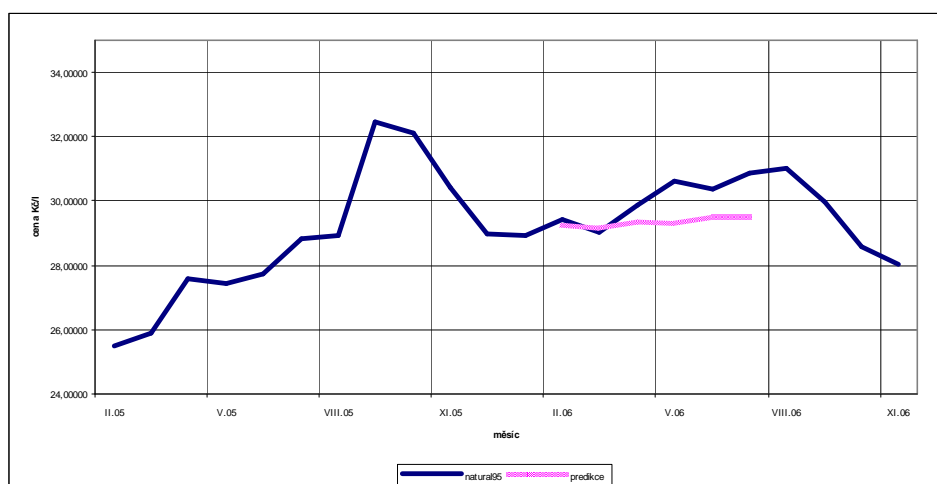
7.1. ARIMA (p, d, q)

V programu SAS byla provedena předpověď modelem, jehož zjednodušený zápis je následující:

$$Y_t = a + b y_{t-1} + c y_{t-2} + d e_{t-1} + v_t. \quad (7.2)$$

Protože jde o ex post prognózu, předpovídané hodnoty lze porovnat se skutečností v **grafu č. 9**.

Graf č. 9: Ex post prognóza ceny benzínu Natural 95 v Kč/l modelem ARIMA (2, 1, 1)



Zdroj: vlastní výpočty

Bodová předpověď relativně vystihuje vývoj ceny benzínu v sezonně očištěné časové řadě, na krátkodobé výkyvy na trhu však zareagovat nemůže.

Statistická verifikace

Výsledný panel odhadnutých parametrů zobrazuje **tabulka č. 4**.

Tabulka č. 4: Panel výsledků u zvolené funkce

Model Parameter	Estimate	Std. Error	T	Prob> T
Intercept	0.06996	0.0790	0.8852	0.3777
Moving Average, Lag 1	-0.89130	0.1386	-6.4323	<.0001
Autoregressive, Lag 1	-0.65369	0.1667	-3.9217	0.0001
Autoregressive, Lag 2	0.09497	0.1088	0.8733	0.3842

Zdroj: vlastní výpočty

- Parametr u a c u funkce (7.2) vyšel statisticky nevýznamný, ostatní jsou významné.
- Residuální složka má charakter bílého šumu, který zachycuje korelogram v **příloze č. 3**.
- $SSE = \sum e^2 = 4,53$

7.2. ADL model

ADL je neomezený dynamický model (jednorovnicový), je zpravidla formulován, respektive definován ve formě ADL (autoregressive distributed lag) modelu.

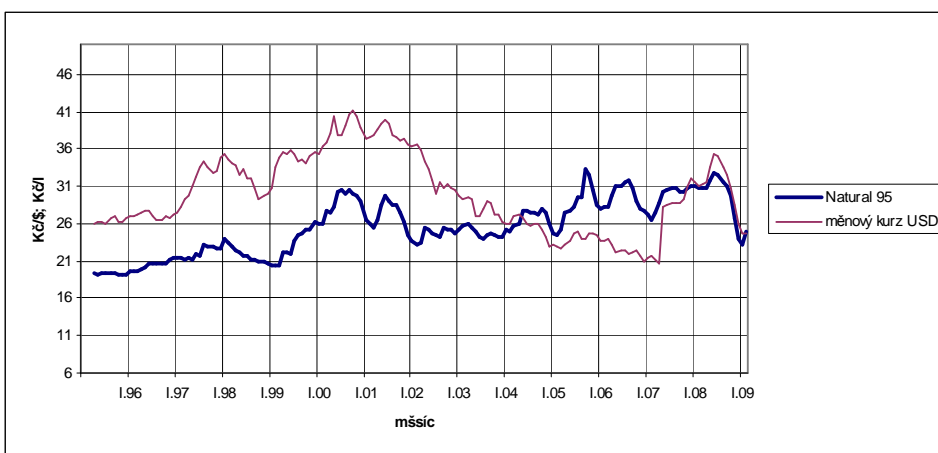
Ceny benzínu budou vysvětlovány průběhem měnového kurzu amerického dolaru. Obecná podoba ADL (n, p) modelu je následující:

$$y_t = \beta_0 + \beta_1 y_{t-1} + \dots + \beta_n y_{t-n} + \gamma_{11} x_{1t-1} + \dots + \gamma_{1p} x_{1t-p} + u_t, \quad (7.3)$$

kde β_0, \dots, β_n a $\gamma_{11}, \dots, \gamma_{1p}$ jsou neznámé parametry pro n zpožděných hodnot endogenní proměnné, p zpoždění exogenní proměnné a u_t je náhodná složka s nulovým podmíněným průměrem, tj. $E(u_t | y_{t-1}, \dots, y_{t-n}, x_{1t}, x_{1t-1}, \dots, x_{1t-p}) = 0$.

Průběh skutečných cen a vybrané exogenní proměnné ukazuje následující **graf č. 10**.

Graf č. 10: Průběh zvolených časových řad



Zdroj: vlastní výpočty

Z uvedeného grafu je vidět, že kurz koruny vůči americkému dolaru nebude pravděpodobně dobrým prediktorem ceny benzínu.

Požadavky na ADL model

- $E(u_t | y_{t-1}, \dots, y_{t-n}, x_{1t}, x_{1t-1}, \dots, x_{1t-p}, \dots, x_{kt}, x_{kt-1}, \dots, x_{kt-p}) = 0$,
- (a) náhodné proměnné ($y_t, x_{1t}, \dots, x_{kt}$) jsou stacionární,
- (b) ($y_t, x_{1t}, \dots, x_{kt}$) a ($y_{t-j}, x_{1t-j}, \dots, x_{kt-j}$) jsou nezávislé s dostatečně velkým j ,
- x_{1t}, \dots, x_{kt} a y_t mají nenulové a konečné první čtyři momenty,
- nepřítomnost perfektní multikolinearity.

Stacionarita náhodných poměňůch je základní požadavek kladených na ADL model, a proto bude otestována. Protože máme pouze jednu exogenní proměnnou, multikolinearitu testovat nemusíme. Ostatní kritéria budeme předpokládat.

Určení řádu integrace

Pro určení stacionarity používáme Dickeyův-Fullerův test, dále jen DF test. Vycházíme ze vztahu:

$$\Delta y_t = \alpha + \delta y_{t-1} + u_t \quad (7.4)$$

Nulová a alternativní hypotéza DF testu má podobu:

- $H_0: \delta = 0$ – Řada je nestacionární a je integrována řádu 1 nebo vyššího a nebo integrována není.
- $H_A: \delta < 0$. – Proces je stacionární.

Vypočtená hodnota t je porovnávána s tabulkovou hodnotou DF testu ($= 2$). Pokud platí, že vypočtená hodnota $|t| > t_{TAB}$, pak zamítáme H_0 a proměnné jsou stacionární, v opačném případě testujeme první diference.

Tabulka č. 5: Panel výsledků u testování stacionarity pro Y_t

Výsledky regrese se závislou proměnnou : Dyt (Tabulka6)					
R= ,15192815 R2= ,02308216 Upravené R2= ,01716145					
F(1,165)=3,8985 p<,05000 Směrod. chyba odhadu : ,89262					
Beta	Sm.chyba beta	B	Sm.chyba B	t(165)	Úroveň p
		0,948323	0,467972	2,02645	0,044329
-0,151928	0,076946	-0,035772	0,018117	-1,97447	0,049997

Zdroj: vlastní výpočty

t-hodnota $|1,97| < 2 = H_0$

Tabulka č. 6: Panel výsledků u testování stacionarity pro Y_t (první diference)

Výsledky regrese se závislou proměnnou : DDYt (Tabulka6)					
R= ,55142845 R2= ,30407333 Upravené R2= ,29985560					
F(1,165)=72,094 p<,00000 Směrod. chyba odhadu : ,83735					
Beta	Sm.chyba beta	B	Sm.chyba B	t(165)	Úroveň p
		0,025604	0,064818	0,39502	0,693341
-0,551428	0,064944	-0,620662	0,073098	-8,49082	0,000000

Zdroj: vlastní výpočty

t-hodnota $|8,48| > 2 = H_a$

Tabulka č. 7: Panel výsledků u testování stacionarity pro X_t

Výsledky regrese se závislou proměnnou : dX_t (ADL_projekt.sta)						
R= ,09599496 R2= ,00921503 Upravené R2= ,00321028						
F(1,165)=1,5346 p<,21718 Směrod. chyba odhadu : 1,0666						
N=167	Beta	Sm.chyba beta	B	Sm.chyba B	t(165)	Úroveň p
Abs.člen			0,586493	0,487546	1,20295	0,230719
X_{t-1}	-0,095995	0,077490	-0,019819	0,015999	-1,23880	0,217180

Zdroj: vlastní výpočty

t-hodnota /1,23/ < 2 = H_0

Tabulka č. 8: Panel výsledků u testování stacionarity pro X_t (první diference)

Výsledky regrese se závislou proměnnou : ddX_t (ADL_projekt.sta)						
R= ,61165166 R2= ,37411775 Upravené R2= ,37032452						
F(1,165)=98,628 p<,00000 Směrod. chyba odhadu : 1,0358						
N=167	Beta	Sm.chyba beta	B	Sm.chyba B	t(165)	Úroveň p
Abs.člen			-0,004685	0,080159	-0,05844	0,953466
Dx_{t-1}	-0,611652	0,061589	-0,744693	0,074986	-9,93116	0,000000

Zdroj: vlastní výpočty

t-hodnota /9,93/ > 2 = H_a

Tabulka č. 9: Panel výsledků u testování stacionarity pro X_t a Y_t

Y_t Cena benzínu Natural 95	-1,97	$\Delta Y_t = \alpha + \delta y_{t-1} + \varepsilon_t$
1. diference	-8,49	$\Delta \Delta Y_t = \alpha + \delta \Delta y_{t-1} + \varepsilon_t$
X_t Kurz USD	-1,23	$\Delta X_t = \alpha + \delta x_{t-1} + \varepsilon_t$
1. diference	-9,93	$\Delta \Delta X_t = \alpha + \delta \Delta x_{t-1} + \varepsilon_t$
t-hodnota	2	

Zdroj: vlastní výpočty

Z výše uvedených tabulek vyplývá, že časové řady jsou tedy integrovány řádu jedna $I(1)$.

Volba délky zpoždění

Volbu délky zpoždění u obou proměnných budeme volit pomocí 3 nejčastěji používaných kritérií:

- R^2 – korigovaného koeficientu determinace,
- AIC (Akaike information criterion),
- BIC (Bayes information criterion),

se vzorci:

$$\bar{R}^2 = 1 - \frac{n-1}{n-q}(1-R^2) \quad (7.5)$$

$$AIC = \ln\left(\frac{SSR}{n} + \frac{2p}{n}\right) \quad (7.6)$$

$$BIC = \ln\left(\frac{SSR}{n}\right) + \frac{p}{n}\ln(n) \quad (7.7)$$

kde SSR je odhadnutá směrodatná odchylka bílého šumu, n je počet pozorování a q počet parametrů v rovnici. Z charakteru vzorců BIC a AIC vyplývá, že budou dávat podobné výsledky. Výstupní tabulky pro volbu délky zpoždění zobrazuje **příloha č. 4**.

Výsledky jednotlivých kritérií potom **tabulka č. 10**.

Tabulka č. 10: Panel výsledků zvolených kritérií

	max_R2	min_AIC	Min_BIC
ADL (1,1)	0,163	-3,191	-5,211
ADL (2,2)	0,172	-2,730	-5,155
ADL (3,3)	0,174	-2,416	-5,094

Zdroj: vlastní výpočty

Pro predikci je vybrán model ADL (1, 1).

Předpověď budoucích hodnot modelem ADL (1, 1)

Pro ex post prognózu byl tedy vybrán model ADL (1, 1), jehož funkce má po odhadu parametrů podobu:

$$y_t = 0,0285 + 0,3245y_{t-1} + 0,1598x_{1t-1} \quad (7.8)$$

Statistická verifikace

- Z tabulky v **příloze č. 4** je vidět, že parametry vyšly statisticky významné.
- Korelogramy reziduální v **příloze č. 5** složky ukazují, že Δe_t splňuje požadavky na bílý šum, a odhadnuté parametry nejsou proto zkresleny.
- $SSE = \sum e^2 = 9,18$.

7.3. VAR model

Model vektorové autoregrese VAR je zobecněním jednorozměrného autoregresního procesu. Vektorové autoregresní modely vycházejí z myšlenky, že všechny proměnné využitě pro analýzu zvolené závislosti jsou náhodné a simultánně závislé. To znamená, že modelová struktura obsahuje pouze endogenní proměnné, přičemž jejich maximální délka zpoždění je stejná.

Další charakteristikou je, že VAR model (jeho obecná forma) nevyhází striktně z ekonomické teorie. Model ADL (1, 1, 1) je upraven na model VAR (1), kde soustava simultánních rovnic v redukovaném tvaru má po odhadu parametrů tvar:

$$y_t = 0,0285 + 0,3245y_{t-1} + 0,1598x_{1t-1}, \quad (7.9)$$

$$y_t = -0,008712 + 0,1715y_{t-1} + 0,223x_{1t-1}, \quad (7.10)$$

kde parametry u rovnice (7.10) zobrazuje **tabulka č. 11**.

Tabulka č. 11: Panel výsledků vybrané funkce

Výsledky regrese se závislou proměnnou : DXt (ADL_KOrekcesta.sta)						
R= ,28319408 R2= ,08019889 Upravené R2= ,06891298						
F(2,163)=7,1061 p<,00110 Směrod. chyba odhadu : 1,0336						
N=166	Beta	Sm.chyba beta	B	Sm.chyba B	t(163)	Úroveň p
Abs. člen			-0,008712	0,080271	-0,108529	0,913709
DYt-1	0,127143	0,077521	0,171364	0,104483	1,640105	0,102912
DXt-1	0,223586	0,077521	0,223509	0,077495	2,884177	0,004455

Zdroj: vlastní výpočty

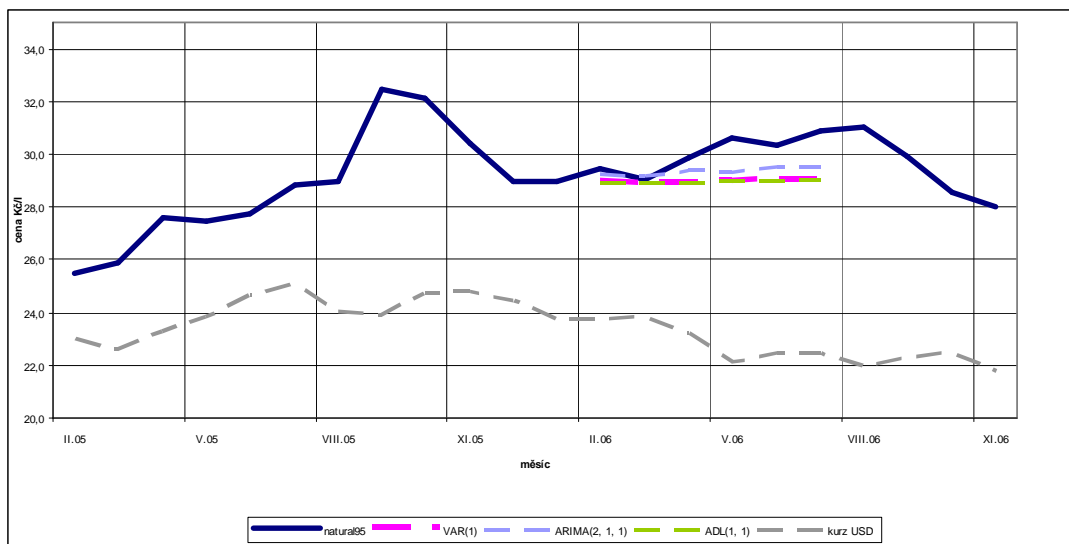
Statistická verifikace

- Z **tabulky č. 11** je vidět, že kromě parametru u proměnné DX_{t-1} , v rovnici (7.10) vyšly parametry statisticky nevýznamné.
- Korelogramy reziduální složky funkce (7.10) zobrazuje **příloha č. 6**. Složka e_t splňuje požadavky na bílý šum, a odhadnuté parametry nejsou proto zkresleny.
- $SSE = \sum e^2 = 9,5$.

Výslednou předpověď všech tří funkcí ukazuje následující **graf č. 11**. Model VAR (1) se liší od ADL (1, 1) tím, že proměnná X_t zde není prognózována průměrem, ale svou vlastní rovnicí. Z vizuální analýzy je vidět, že obě funkce dávají téměř shodné výsledky. Model ARIMA (2, 1, 1) časovou řadu popisuje nejlépe. Důvodem mohou být klesající hodnoty měnového kurzu v době

předpovědi, který je zde jako exogenní proměnná X_{t-1} . Klesající měnový kurz zde snižuje hodnoty \hat{Y}_t , kdy jejich skutečná hodnota roste.

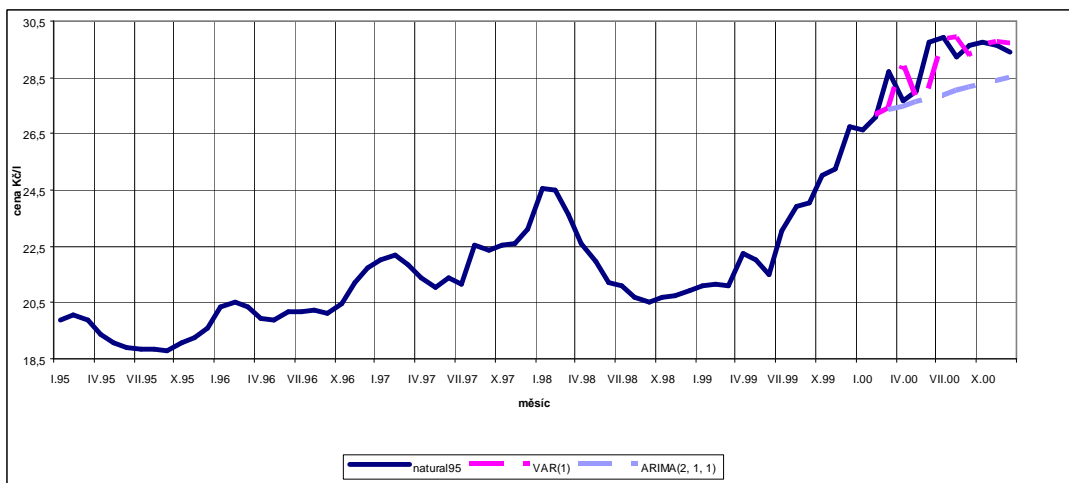
Graf č. 11: Ex post prognóza ceny benzínu Natural 95 v Kč/l (první období)



Zdroj: vlastní výpočty

Z výše uvedených důvodů byla provedena další ex post prognóza na období od března 2000 do června 2000. Tento interval bude dále nazýván jen jako *druhé období* a predikce v **grafu č. 11** jako *období první*. V tomto období je z vizuální analýzy **grafu č. 10** patrné, že změny X_t alespoň přibližně souvisejí se změnou Y_t . **Graf č. 12** zachycuje průběh sezonně očištěných cen benzínu, vyrovnané hodnoty modelem ARIMA (2, 1, 1) a VAR (1).

Graf č. 12: Ex post prognóza ceny benzínu Natural 95 v Kč/l (druhé období)



Zdroj: vlastní výpočty

Průběh změn Y_t lze částečně vysvětlit změnou X_t , a proto prognóza modelem VAR (1) podává mnohem lepší výsledek než model ARIMA. Rozdíl sledovaného kritéria SSE je mezi ADL a VAR modelem pouhých 0,06 ve prospěch modelu ADL. Průběh časových řad by se v jednom grafu překrýval.

8. Závěr

Modely byly zkoumány na sezonně očištěné řadě cen benzínu Natural 95. Kritériem byl součet čtvercových chyb SSE. Následující tabulka ukazuje rozdíl mezi skutečnou a vyrovnanou hodnotou v prvním předpovídaném období a také zvolené kritérium SSE pro prognostický horizont $h = 6$.

Tabulka č. 12: Porovnání vybraných přístupů

	1.období		2. období	
	$e_{t=n+1}^2$	$\sum_{t=n+1}^{t=n+h} e_t^2$	$e_{t=n+1}^2$	$\sum_{t=n+1}^{t=n+h} e_t^2$
ARIMA (2, 1, 1)	0,023	4,76	1,95	11,76
ADL (1, 1)	0,312	9,18	1,66	6,4
VAR (1)	0,312	9,5	1,66	6,46

Zdroj: vlastní výpočty

Prognózováním hodnot X_t (měnový kurz) v **prvním období** nebylo dosaženo lepšího výsledku než u modelu ADL, kde hodnoty této proměnné byly nahrazeny průměrem. Důvodem je opačný průběh zvolených časových řad v době ex post prognózy. Posilující koruna vůči dolaru neustále snižuje vyrovnanou hodnotu ceny benzínu v době, kdy skutečná cena na trhu roste. Tento opačný vývoj obou časových řad má za následek to, že model ADL (1, 1) dává (byť zanedbatelně) lepší výsledek, což v praxi obvykle neplatí. Tento fakt je pravděpodobně způsoben špatně vybraným regresorem X_t . Kurz koruny vůči dolaru zhruba do roku 2002 alespoň přibližně kopíruje vývoj ceny benzínu, od tohoto roku však koruna posiluje a v předpovídaném období kurz stále klesá, navzdory růstu ceny paliva.

Proto byla provedena verifikace pro **druhé období**, kde „ekonometrické“ modely dávaly podstatně lepší výsledek než model u jednorozměrné časové řady. Model VAR (1) však nepodal, z hlediska sledovaného kritéria, lepší výsledek než model ADL (1, 1). Závěrem lze říci, že měnový kurz je jako regresor ceny benzínu nevhodný.

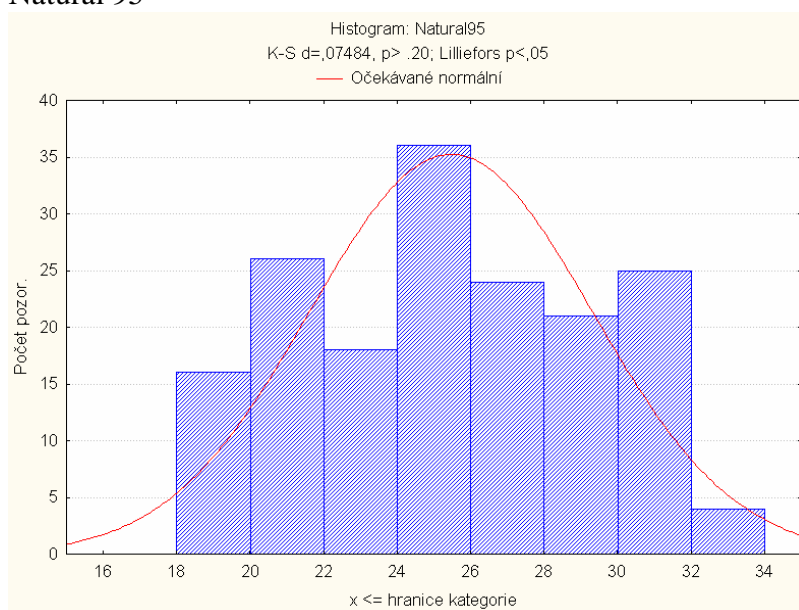
9. Použité zdroje

- CIPRA, Tomáš. *Finanční ekonometrie*. [s.l.] : Ekopress, 2008. 520 s. ISBN 978-80-86929-43-9.
- Ing. Lukáš Čechura, Ph.D. *Přednášky*.
- <http://www.cnb.cz/cs/>
- <http://www.czso.cz/>

10. Přílohy

Příloha č. 1: Histogram ceny benzínu Natural 95

Natural 95



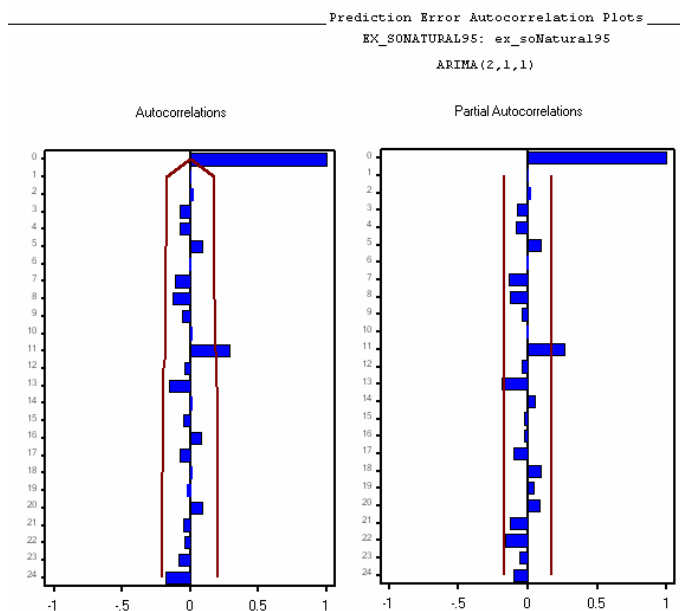
Zdroj: vlastní výpočty

Příloha č. 2: Odhadnuté hodnoty lineární funkce u benzínu Natural 95 v Kč/l

DATE	ACTUAL	PREDICT	U95	L95
01FEB2009	26.0359	30.8359	34.9765	26.6953
01MAR2009	.	30.8993	35.0399	26.7587
01APR2009	.	30.9627	35.1033	26.8221
01MAY2009	.	31.0261	35.1667	26.8856
01JUN2009	.	31.0896	35.2301	26.9490
01JUL2009	.	31.1530	35.2936	27.0124
01AUG2009	.	31.2164	35.3570	27.0758

Zdroj: vlastní výpočty

Příloha č. 3: Korelogramy reziduí modelu ARIMA (2, 1, 1)



Zdroj: vlastní výpočty

Příloha č. 4: Panely výsledků pro volbu délky zpoždění

ADL (1, 1)

Výsledky regrese se závislou proměnnou : DYt (ADL_volbaSpozde						
R= ,41638408 R2= ,17337570 Upravené R2= ,16323307						
F(2,163)=17,094 p<,00000 Směrod. chyba odhadu : ,82610						
N=166	Beta	Sm.chyba beta	B	Sm.chyba B	t(163)	Úroveň p
Abs.člen			0,028541	0,064149	0,444910	0,656975
DYt-1	0,320522	0,074293	0,324588	0,075236	4,314279	0,000028
DXt-1	0,189706	0,074293	0,159886	0,062615	2,553477	0,011582

Zdroj: vlastní výpočty

ADL (2, 2)

Výsledky regrese se závislou proměnnou : DYt (ADL_volbaSpozdeni.						
R= ,43878728 R2= ,19253428 Upravené R2= ,17247302						
F(4,161)=9,5973 p<,00000 Směrod. chyba odhadu : ,82153						
N=166	Beta	Sm.chyba beta	B	Sm.chyba B	t(161)	Úroveň p
Abs.člen			0,032314	0,063823	0,50631	0,613332
DYt-1	0,350713	0,080157	0,355163	0,081174	4,37531	0,000022
DYt-2	-0,141124	0,078577	-0,143176	0,079720	-1,79599	0,074371
DXt-1	0,191022	0,075533	0,160994	0,063660	2,52899	0,012401
DXt-2	0,081257	0,076679	0,068559	0,064697	1,05970	0,290870

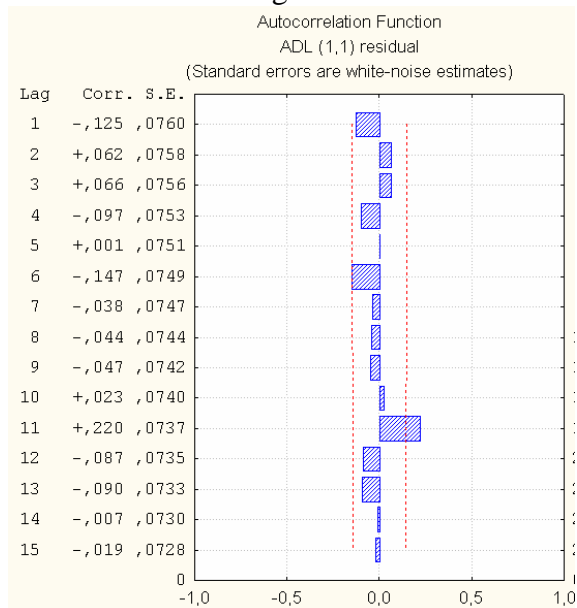
Zdroj: vlastní výpočty

ADL (3, 3)

Výsledky regrese se závislou proměnnou : DYt (ADL_volbaSpozdeni						
R= ,45189359 R2= ,20420782 Upravené R2= ,17417793						
F(6,159)=6,8002 p<,00000 Směrod. chyba odhadu : ,82068						
N=166	Beta	Sm.chyba beta	B	Sm.chyba B	t(159)	Úroveň p
Abs.člen			0,036270	0,063839	0,56814	0,570738
DYt-1	0,346029	0,080988	0,350419	0,082015	4,27260	0,000033
DYt-2	-0,092341	0,084748	-0,093684	0,085980	-1,08960	0,277536
DYt-3	-0,103417	0,078488	-0,108692	0,082491	-1,31763	0,189524
DXt-1	0,195171	0,075880	0,164492	0,063952	2,57211	0,011024
DXt-2	0,099468	0,077923	0,083925	0,065747	1,27648	0,203646
DXt-3	-0,043471	0,076640	-0,037221	0,065622	-0,56720	0,571377

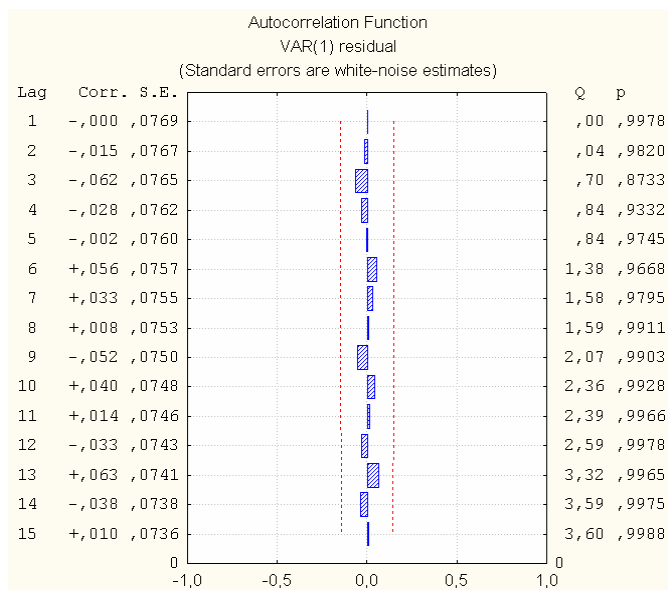
Zdroj: vlastní výpočty

Příloha č. 5: Korelogram autokorelační funkce reziduí modelu ADL (1, 1)



Zdroj: vlastní výpočty

Příloha č. 6: Korelogram autokorelační funkce reziduí modelu VAR (1)



Zdroj: vlastní výpočty