

UNIVERZITA KOMENSKÉHO V BRATISLAVE

FAKULTA MATEMATIKY, FYZIKY A INFOMATIKY



**ANALÝZA KORELÁCIE POHYBOV CIEN ZEMNÉHO
PLYNU, ROPY A KURZU USD/EUR A NÁVRH
TRADINGOVEJ STRATÉGIE PRE OBCHODNÍKA
S PLYNOM NA ZMENY JEDNOTLIVÝCH PARAMETROV**

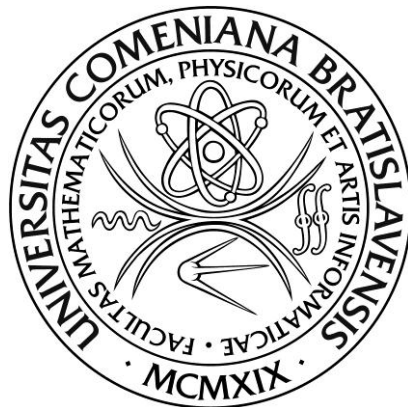
Diplomová práca

Bratislava, 2013

Bc. Alexander Róža

UNIVERZITA KOMENSKÉHO V BRATISLAVE

FAKULTA MATEMATIKY, FYZIKY A INFOMATIKY



**ANALÝZA KORELÁCIE POHYBOV CIEN ZEMNÉHO
PLYNU, ROPY A KURZU USD/EUR A NÁVRH
TRADINGOVEJ STRATÉGIE PRE OBCHODNÍKA
S PLYNOM NA ZMENY JEDNOTLIVÝCH PARAMETROV**

Diplomová práca

Študijný program: Ekonomická a finančná matematika
Študijný odbor: 9.1.9. Aplikovaná matematika, 1114
Školiace pracovisko: Katedra aplikovanej matematiky a štatistiky,
FMFI UK, Mlynská dolina, 842 48 Bratislava
Školiteľ: Mgr. Matej Krušpán

Bratislava, 2013

Bc. Alexander Róža



Univerzita Komenského v Bratislave
Fakulta matematiky, fyziky a informatiky

ZADANIE ZÁVEREČNEJ PRÁCE

Meno a priezvisko študenta: Bc. Alexander Róža
Študijný program: ekonomická a finančná matematika (Jednoodborové štúdium, magisterský II. st., denná forma)
Študijný odbor: 9.1.9. aplikovaná matematika
Typ záverečnej práce: diplomová
Jazyk záverečnej práce: slovenský

Názov: Analýza korelácie pohybov cien zemného plynu, ropy a kurzu USD/EUR a návrh tradingovej stratégie pre reakcie obchodníka s plynom na zmeny jednotlivých parametrov

Cieľ: Cieľom diplomovej práce je určiť koreláciu cien v horizonte deň, týždeň, mesiac a podľa toho navrhnúť stratégiu reakcie obchodníka na hube v EU tak, aby výsledná pozorovaná cena bola blízka prognózovanej v danom časovom úseku. Predpokládame pritom, že z dlhodobého hľadiska (1 mesiac) sa cena plynu vždy ustáli na rovnovážnej cene podľa korelácie s ropou a FX ako jedinými fundamentálnymi cenovými faktormi.

Vedúci: Mgr. Matej Krušpán
Katedra: FMFI.KAMŠ - Katedra aplikovanej matematiky a štatistiky
Vedúci katedry: prof. RNDr. Daniel Ševčovič, CSc.
Dátum zadania: 25.01.2012

Dátum schválenia: 26.01.2012

prof. RNDr. Daniel Ševčovič, CSc.
garant študijného programu

.....
študent

.....
vedúci práce

Čestné prehlásenie

Čestne prehlasujem, že som túto diplomovú prácu vypracoval samostatne pod dohľadom vedúceho diplomovej práce a s použitím uvedenej literatúry.

V Bratislave, apríl 2013

.....
Bc. Alexander Róža

Pod'akovanie

Touto cestou by som sa rád pod'akoval vedúcemu diplomovej práce Mgr.Matejovi Krušpánovi za venovaný čas, rady a motiváciu pri písaní. Zároveň by som sa rád pod'akoval Prof.Dr.Jarkovi Fidrmucovi za cenné rady pri tvorbe modelov.V neposlednom rade d'akujem rodine a priateľke za podporu počas štúdia.

RÓŽA, Alexander: Analýza korelácie pohybov cien zemného plynu, ropy a kurzu USD/EUR a návrh tradingovej stratégie pre reakcie obchodníka s plynom na zmeny jednotlivých parametrov [Diplomová práca], Univerzita Komenského v Bratislave; Fakulta matematiky, fyziky a informatiky; Katedra aplikovanej matematiky a štatistiky. Vedúci diplomovej práce: Mgr. Matej Krušpán, Bratislava, 2013, 46s.

Abstrakt

Táto diplomová práca sa venuje využitiu lineárnych časových radov na zostavenie predikčného modelu pre prognózu cien zemného plynu na huboch v Európe. Na začiatku sa venujeme teoretickej časti, špecifikácii lineárnych modelov. Definujeme si klasický lineárny regresný model a všetky predpoklady, ktoré by mal spĺňať. V praktickej časti aplikujeme teoretické poznatky a správnosť skonštruovaných modelov overíme porovnaním predikovaných hodnôt so skutočnými pozorovaniami.

Kľúčové slová: časový rad, ARIMA, VAR, cena plynu, hub, Grangerova kauzalita

RÓŽA, Alexander: Analysis of correlation of natural gas price movements, crude oil price and USD/EUR exchange rate and designing a trading strategy for response on changes in individual parameters.[Master's thesis], Comenius University in Bratislava; Faculty of Mathematics, Physics and Informatics; Department of Applied Mathematics and Statistics. Master thesis supervisor: Mgr.Matej Krušpán, Bratislava, 2013, 46p.

Abstract

The goal of this thesis is to build a prediction model to forecast natural gas prices on gas hubs in Europe using linear time series. At the beginning we will talk about theory and specify a linear model. We define a classical linear regression model and all his assumptions. In the practical part we apply the theoretical knowledge and validate the engineered models by comparing the predicted values with actual observations.

Key words: time series, ARIMA, VAR, natural gas price, gas hub, Granger causality

Obsah

Úvod.....	2
1. Základné pojmy.....	3
1.1 Stochastický proces a časový rad.....	3
1.2 Stacionarita procesu.....	5
1.3 Autokorelačná a parciálna autokorelačná funkcia.....	6
1.3.1 Autokorelačná funkcia.....	6
1.3.2 Parciálna autokorelačná funkcia.....	6
1.4 Biely šum.....	8
1.5 Operátor spätného posunutia.....	8
2. Jednorovnicové lineárne modely.....	10
2.1 $AR(p)$ proces.....	10
2.2 $MA(q)$ proces.....	11
2.3 $ARMA(p, q)$ proces.....	11
2.4 $ARIMA(p, d, q)$ proces.....	12
3. Vektorovo autoregresné modely.....	13
3.1 VAR modely – Štrukturálny a redukovaný tvar.....	13
3.2 Stabilita.....	16
3.3 Woldova dekompozícia.....	17
3.4 Impulzné odozvy (Impulse response).....	17
3.5 Grangerova kauzalita.....	18
4. Model lineárnej regresie.....	20
4.1 Identifikácia modelu.....	20
4.2 Odhad parametrov.....	20
4.3 Verifikácia modelu.....	21
4.3.1 Stacionarita – Dickey-Fullerov test.....	21
4.3.2 Koeficient determinácie.....	22
4.3.3 Heteroskedasticita.....	23
4.3.4 Autokorelovanosť reziduí.....	23
4.3.5 Informačné kritériá pre výber modelu.....	24
5. Praktická časť.....	25
5.1 ARIMA modelovanie.....	25
5.2 VAR modelovanie.....	39

5.3 Grangerova kauzalita	43
5.4 Impulse response.....	44
Záver	45
Zoznam použitej literatúry	46

Úvod

V ekonomickej praxi sa veľmi často stretávame so súbormi pozorovaní, ktoré sú usporiadané v čase, tým pádom ich môžeme považovať za časové rady. Analyzovaním správania týchto pozorovaní sa snažíme predpovedať, ako by sa mohli vyvíjať v budúcnosti a v dôsledku tohto poznania docieľiť prípadný profit. V tejto diplomovej práci sa zameriame práve na tvorbu správnych modelov, ktoré nám pomôžu predpovedať budúci vývoj sledovaných pozorovaní.

Práca pozostáva z piatich kapitol. V prvej kapitole si zdefinujeme základné štatistické a ekonometrické pojmy, čím čitateľovi priblížime teoretické poznatky z oblasti časových radov. Vysvetlíme si, čo rozumieme pod pojmi stochastický proces, stacionarita, autokorelačná a parciálna autokorelačná funkcia, biely šum.

V druhej kapitole si pomocou Box-Jenkinsovej metodológie predstavíme jednorovnicové lineárne AR, MA, ARMA a ARIMA modely.

V tretej kapitole sa budeme zaoberať v ekonómii populárnym vektorovo autoregresným modelom. Uvedieme štrukturálny a redukovaný tvar VAR modelu, objasníme podmienky jeho stability. Ďalej sa budeme zaoberať Woldovou dekompozíciou, ktorá nás privedie k impulzným odozvám a Grangerovej kauzalite.

V poslednej teoretickej kapitole bude predmetom skúmania klasický model lineárnej regresie, zameriame sa na správnu identifikáciu modelu. Predstavíme si metódu odhadu parametrov, ktorú budeme používať v praktickej časti. Následne sa zameriame na verifikáciu zvoleného modelu, kde si uvedieme základné charakteristiky, ako sú stacionarita, koeficient determinácie, heteroskedasticita a autokorelácia reziduí, ktoré budeme overovať pre správne určenie modelu.

V piatej, praktickej časti uvedieme do praxe teoretické poznatky z predchádzajúcich kapitol. Aplikujeme ich na dáta denných cien zemného plynu na obchodných trhoch (HUBoch) v Európe. Budeme sa snažiť vytvoriť čo najpresnejšie modely pre prognózu cien s využitím vzájomnej korelácie týchto trhov.

Výsledky nášho výskumu budú zhrnuté v závere tejto práce.

Kapitola 1

Základné pojmy

V tejto kapitole si uvedieme definície niektorých pojmov zo štatistiky, náhodných procesov, ekonometrických modelov a iných oblastí matematiky, na ktoré budeme v nasledujúcich kapitolách odkazovať.

1.1 Stochastický proces a časový rad

Definícia 1.1

Trojicu (Ω, S, P) nazývame *pravdepodobnostný priestor*, pričom Ω je priestor elementárnych udalostí ω , S je σ -algebra podmnožín priestoru elementárnych udalostí. Prvkom z S priradíme určitú pravdepodobnosť pomocou pravdepodobnostnej miery P .

Definícia 1.2

Nech je daný pravdepodobnostný priestor (Ω, S, P) . *Náhodným vektorom* (n -rozmernou náhodnou veličinou) $X = (X_1, X_2, \dots, X_n)^T$ nazývame zobrazenie

$$X: \Omega \rightarrow R^n; \forall x = (x_1, x_2, \dots, x_n)^T \in R^n : \bigcap_{i=1}^n \{\omega, X_i(\omega) < x_i\} \in S$$

Definícia 1.3

Nech máme daný pravdepodobnostný priestor (Ω, S, P) , indexovú množinu $T \subseteq R$ a reálnu funkciu $X: \Omega \times T \rightarrow R$ definovanú pre $\forall \omega \in \Omega$ a $\forall t \in T$. Ak je $X(\omega, t)$ pre $\forall t \in T$ náhodá veličina, potom funkciu $X(\omega, t)$ nazývame *náhodný proces*. Značíme $\{X_t, t \in T\}$.

Podľa indexovej množiny času T rozlišujeme dva typy náhodných procesov a to

- náhodné procesy s diskretným časom
- náhodné procesy so spojitým časom

Pre realizáciu stochastického procesu s diskretným časom budeme používať pojem *časový rad*. V ďalšom texte preto budeme uvažovať len stochastické procesy s diskretným časom.

Pre každý stochastický proces platia nasledovne definované charakteristiky:

Definícia 1.4

Ak existuje pre $\forall t \in T$ stredná hodnota $E(X_t)$, tak funkciu

$$\mu_t = E(X_t)$$

nazývame strednou hodnotou stochastického procesu.

Definícia 1.5

Nech $\{X_t, t \in T\}$ je stochastický proces, ktorý má konečné druhé momenty, tj. $E(X_t^2) < \infty$ pre $\forall t \in T$. Potom funkciu

$$\gamma(s, t) = cov(s, t) = E[(X_s - \mu_s)(X_t - \mu_t)]$$

nazývame *autokovariančnou funkciou*.

Kovariancia meria lineárnu závislosť medzi náhodnými veličinami. Ak nastane prípad, že v definícii 1.5 $s = t$, hovoríme o *variančnej funkcii*

$$\sigma_t^2 = D(X_t) = E(X_t - \mu_t)^2$$

Definícia 1.6

Nech T^n je množina všetkých vektorov

$$T^n = \{t = (t_1, \dots, t_n)': t_1 \leq t_2 \leq \dots \leq t_n; t_i \in T; i = 1, \dots, n\}.$$

Potom *distribučnou funkciou* stochastického procesu rozumieme funkciu

$$F_t(x) = F_{t_1, \dots, t_n}(x_1, \dots, x_n) = P(X_{t_1} \leq x_1, \dots, X_{t_n} \leq x_n)$$

pre $\forall t = (t_1, \dots, t_n)' \in T^n$ a $\forall x = (x_1, \dots, x_n)' \in R^n$.

1.2 Stacionarita procesu

Teraz si postupne zdefinujeme jednotlivé druhy stacionarity, ktoré rozlišujeme v teórii stochastických procesov.

Definícia 1.7

Stochastický proces $\{X_t, t \in T\}$ je *striktne stacionárny*, ak pre $\forall t = (t_1, \dots, t_n) \in T^n$ a pre $\forall \tau = (\tau_1, \dots, \tau_n) \in T^n$ platí

$$F_t(x) = F_{t_1, \dots, t_n}(x_1, \dots, x_n) = F_{\tau_1, \dots, \tau_n}(x_1, \dots, x_n) = F_\tau(x)$$

Túto definíciu môžeme interpretovať tak, že bez ohľadu v akom čase sa nachádzame v rámci časového radu, vektor prvkov časového radu rovnakej dĺžky musí mať rovnaké pravdepodobnostné rozdelenie.

Definícia 1.8

Stochastický proces $\{X_t, t \in T\}$ je *stacionárny v strednej hodnote* práve vtedy, keď pre $\forall t \in T$ je stredná hodnota konštantná.

$$E(X_t) = \mu$$

Definícia 1.9

Stochastický proces $\{X_t, t \in T\}$ je *kovariančne stacionárny*, ak pre $\forall t, s \in T$ platí

$$\gamma(s, t) = \gamma(s - t)$$

Teda kovariančná funkcia je len funkciou $(s - t)$, čiže je invairantná voči posunom v čase a závisí iba od vzdialenosti pozorovaní.

Z definície nám ďalej vyplýva

$$D(X_t) = \text{Cov}(X_t, X_t) = \gamma(t - t) = \gamma(0).$$

Definícia 1.10

Stochastický proces $\{X_t, t \in T\}$ sa nazýva (slabo) *stacionárny*, ak je kovariančne stacionárny a stacionárny v strednej hodnote.

Namiesto označenia slabo stacionárny proces budeme v ďalšom texte používať pojem *stacionárny proces*.

1.3 Autokorelačná a parciálna autokorelačná funkcia

Pre vhodnú identifikáciu modelu pre časový rad zohrávajú dôležitú úlohu autokorelačná funkcia a parciálna autokorelačná funkcia.

1.3.1 Autokorelačná funkcia

Nech stochastický proces $\{X_t, t \in T\}$ je stacionárny. Potom funkcia

$$\rho(k) = \frac{\text{Cov}(X_t, X_{t-k})}{\sqrt{D(X_t)}\sqrt{D(X_{t-k})}} = \frac{\gamma(k)}{\gamma(0)}$$

sa nazýva *autokorelačná funkcia* stacionárneho procesu, kde vzhľadom k stacionarite procesu $D(X_t) = D(X_{t-k}) = \gamma_0$. Autokorelačná funkcia sa označuje ako ACF.

V prípade stacionárneho stochastického procesu má autokorelačná funkcia nasledujúce vlastnosti

- $\rho_0 = 1$
- $|\gamma_k| \leq \gamma_0 ; |\rho_k| \leq 1$
- $\gamma_k = \gamma_{-k}$ a $\rho_k = \rho_{-k}$ pre všetky k

Autokorelačná funkcia je teda symetrická okolo k , nadobúda hodnoty z intervalu $(-1,1)$ a jej graf sa nazýva korelogram.

1.3.2 Parciálna autokorelačná funkcia

Korelácia medzi dvoma náhodnými veličinami je často spôsobená tým, že obe tieto veličiny sú korelované s treťou veličinou. Značná časť korelácie medzi veličinami X_t a X_{t-k} môže byť teda spôsobená ich koreláciou s veličinami $X_1, X_2, \dots, X_{t-k+1}$. Parciálne autokorelácie nám poskytujú informáciu o korelácii veličín X_t a X_{t-k} , ktorá je očistená o vplyv veličín, ktoré ležia medzi nimi.

Parciálnu autokoreláciu s oneskorením k vyjadruje parciálny regresný koeficient ϕ_{kk} v autoregresii k -teho rádu

$$X_t = \phi_{k1}X_{t-1} + \phi_{k2}X_{t-2} + \dots + \phi_{kk}X_{t-k} + e_t,$$

kde veličina e_t je nekorelovaná s veličinami $X_{t-j}, j \geq 1$. Táto matica je vždy regulárna, takže môžeme použiť Cramerovo pravidlo. Ak vynásobíme obe strany rovnice veličinou X_{t-j} potom stredná hodnota tejto rovnice nadobudne tvar

$$\gamma_j = \phi_{k1}\rho_{j-1} + \phi_{k2}\rho_{j-2} + \dots + \phi_{kk}\rho_{j-k}$$

takže platí

$$\rho_j = \phi_{k1}\rho_{j-1} + \phi_{k2}\rho_{j-2} + \dots + \phi_{kk}\rho_{j-k}$$

Pre $j = 1, 2, \dots, k$ dostávame nasledujúci systém rovníc

$$\rho_1 = \phi_{k1}\rho_0 + \phi_{k2}\rho_1 + \dots + \phi_{kk}\rho_{k-1}$$

$$\rho_2 = \phi_{k1}\rho_1 + \phi_{k2}\rho_0 + \dots + \phi_{kk}\rho_{k-2}$$

...

$$\rho_k = \phi_{k1}\rho_{k-1} + \phi_{k2}\rho_{k-2} + \dots + \phi_{kk}\rho_0$$

Postupným použitím Cramerovho pravidla pre $k = 1, 2, \dots$ dostávame

$$\phi_{11} = \rho_1,$$

$$\phi_{22} = \frac{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & \rho_2 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix}} = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2},$$

⋮

$$\phi_{kk} = \frac{\begin{vmatrix} 1 & \rho_1 & \rho_2 & \dots & \rho_{k-2} & \rho_1 \\ \rho_1 & 1 & \rho_1 & \dots & \rho_{k-3} & \rho_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \dots & \rho_1 & \rho_k \\ 1 & \rho_1 & \rho_2 & \dots & \rho_{k-2} & \rho_{k-1} \\ \rho_1 & 1 & \rho_1 & \dots & \rho_{k-3} & \rho_{k-2} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \dots & \rho_1 & 1 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \rho_2 & \dots & \rho_{k-2} & \rho_1 \\ \rho_1 & 1 & \rho_1 & \dots & \rho_{k-3} & \rho_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \dots & \rho_1 & \rho_k \\ 1 & \rho_1 & \rho_2 & \dots & \rho_{k-2} & \rho_{k-1} \\ \rho_1 & 1 & \rho_1 & \dots & \rho_{k-3} & \rho_{k-2} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \dots & \rho_1 & 1 \end{vmatrix}}$$

V literatúre sa parciálna autokorelácia označuje ako ϕ_{kk} . Ako funkcia oneskorenia k sa ϕ_{kk} nazýva *parciálna autokorelačná funkcia* – PACF.

1.4 Biely šum

Definícia 1.11

Stochastický proces $\{\varepsilon_t, \forall t \in T\}$, kde ε_t sú nekorelované náhodné veličiny so strednou hodnotou $E(\varepsilon_t) = 0$ a rozptylom $D(\varepsilon_t) = \sigma_\varepsilon^2$ nazývame biely šum. Značíme ho $\varepsilon_t \sim WN(0, \sigma_\varepsilon^2)$.

Z tejto definície vyplíva, že proces bieleho šumu je centrováný stacionárny stochastický proces a s autokovariančnou funkciou $\gamma(k) = 0$ pre $k \neq 0$, autokorelačnou funkciou

$$\rho(k) = \begin{cases} 1, & k = 0 \\ 0, & k \neq 0 \end{cases}$$

a parciálnou autokorelačnou funkciou

$$\phi_{kk}(k) = \begin{cases} 1, & k = 0 \\ 0, & k \neq 0 \end{cases}$$

V prípade, že v definícii 1.11 sú nezávislé ε_t , jedná sa o IID proces a značíme ho $\varepsilon_t \sim IID(0, \sigma_\varepsilon^2)$. Ak je zdužené rozdelenie pravdepodobnosti bieleho šumu normálne, hovoríme, že je gaussovský.

1.5 Operátor spätného posunutia

Definícia 1.12

Nech $\{X_t, t \in T\}$ je stochastický proces. Pre operátor spätného posunutia B platí

$$BX_t = X_{t-1}$$

Tento operátor môžeme aplikovať niekoľkokrát za sebou. Platí

$$B^j X_t = X_{t-j}$$

1.6 Diferenčný operátor

Definícia 1.13

Nech $\{X_t, t \in T\}$ je stochastický proces. Pre diferenčný operátor Δ platí

$$\Delta X_t = X_t - X_{t-1} = (1 - B)X_t$$

Tento operátor môžeme takisto aplikovať niekoľkokrát za sebou, pričom platí

$$\Delta^j X_t = (1 - B)^j X_t.$$

Kapitola 2

Jednorovnicové lineárne modely

V tejto kapitole si predstavíme základné modely Box-Jenkinsovej metodológie, ktorá bola vytvorená v sedemdesiatych rokoch 20.storočia. Pochopenie týchto modelov je dôležitým predpokladom pre pochopenie zložitejších modelov, ktoré budeme opisovať v kapitole 3.

2.1 $AR(p)$ proces

Autoregresný proces rádu p vieme zapísať nasledovne

$$X_t = \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + \varepsilon_t$$

Ak je stredná hodnota procesu μ rôzna od nuly, potom proces $AR(p)$ nadobúda tvar

$$X_t = \phi_0 + \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + \varepsilon_t$$

pričom $\phi_0 = \mu(1 - \sum_{j=1}^p \phi_j)$. V operátorovom tvare sa dá proces $AR(p)$ zapísať nasledovne

$$\Phi(B)X_t = \varepsilon_t,$$

kde

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

Takýto proces je *stacionárny* vtedy, keď korene polynomiálnej rovnice

$$(1 - \phi_1 B - \dots - \phi_p B^p) = 0$$

ležia mimo jednotkového kruhu. ACF takéhoto procesu tvorí kombinácia exponenciálne klesajúcich pohybov (v prípade reálnych koreňov $\Phi(B)$) a exponenciálne klesajúcich sínusoidných pohybov (v prípade komplexných koreňov $\Phi(B)$). PACF má pre oneskorenie $k = 1, 2, \dots, p$ hodnoty rôzne od nuly a pre ďalšie oneskorenie sú tieto hodnoty nulové.

2.2 MA(q) proces

Proces kĺzavých priemerov (moving average) rádu q má tvar

$$X_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_p \varepsilon_{t-p}$$

Ak je stredná hodnota procesu μ rôzna od nuly, potom proces $MA(q)$ nadobúda tvar

$$X_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_p \varepsilon_{t-p}$$

Pomocou operátora spätného pounutia môžeme tento proces zapísať ako

$$X_t = \theta(B)\varepsilon_t,$$

kde

$$\theta(B) = (1 + \theta_1 B + \dots + \theta_q B^q).$$

Definícia 2.1

Ak existuje absolútne konvergentná postupnosť reálnych čísel $\{\pi_j\}_{j=0}^{\infty}$ taká, že

$$\varepsilon_t = \sum_{j=0}^{\infty} \pi_j X_{t-j}$$

tak $MA(q)$ proces sa nazýva *invertibilný*.

Invertibilita sa dá zapísať podmienkou

$$\theta(z) = (1 + \theta_1 z + \dots + \theta_q z^q) \neq 0 \text{ pre } |z| \leq 0.$$

Teda všetky korene polynómu $\theta(z)$ musia ležať mimo jednotkového kruhu. ACF takéhoto procesu má pre oneskorenie $k > q$ nulové hodnoty. PACF je tvorená kombináciou exponenciálne klesajúcivh sínusoidných pohybov a exponenciálne klesajúcich pohybov podobne ako v prípade ACF v procese $AR(P)$.

2.3 ARMA(p, q) proces

Tento proces vieme zapísať nasledovne

$$X_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p} = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_p \varepsilon_{t-p}$$

a pomocou operátora spätného posunutia ako

$$\Phi(B)X_t = \theta(B)\varepsilon_t$$

Ak $\mu \neq 0$, potom tento proces nadobúda tvar

$$\Phi(B)X_t = \mu \left(1 - \sum_{j=1}^p \phi_j B^j \right) + \theta(B)\varepsilon_t.$$

Tvar ACF a PACF exponenciálne alebo oscilačne klesá k nule od oneskorenia p resp. q . Aby bol proces $ARMA(p, q)$ stacionárny, korene rovnice $\Phi(B) = 0$ musia ležať mimo jednotkového kruhu. Aby bol proces $ARMA(p, q)$ invertibilný, korene rovnice $\theta(B) = 0$ musia takisto ležať mimo jednotkového kruhu.

2.4 $ARIMA(p, d, q)$ proces

Integrovaný autoregresný proces kľzavých priemerov rádu p, d, q je definovaný v operátorovom tvare ako

$$\Phi(B)\Delta^d X_t = \theta(B)\varepsilon_t$$

Tento proces je nestacionárny a stáva sa stacionárnym diferencovaním rádu d ,

$$W_t = \Delta^d X_t,$$

pričom W_t je stacionárny $ARMA(p, q)$.

Kapitola 3

Vektorovo autoregresné modely

V ekonomickej praxi často potrebujeme zachytiť vzájomné väzby medzi rôznymi premennými. Toto sa nedá docieľiť pomocou jednorovnicových modelov, pretože sú v nich vzťahy zredukované len na závislosť jednej vysvetľovanej premennej na niekoľko vysvetľujúcich premenných a na náhodnú zložku.

Pokiaľ máme sústavu rovníc, kde aspoň jedna rovnica obsahuje viac endogénnych premenných a celkový počet endogénnych premenných je rovný počtu rovníc, ide o model simultánnych rovníc. Ak tieto rovnice obsahujú aj rôzne minulé hodnoty zahrnutých premenných, má tento systém dynamický charakter.

Problémom je, že ekonomická teória nemusí pri špecifikácii dynamického modelu simultánnych rovníc poskytovať dostatočné informácie pre jednoznačný popis štruktúry, napríklad je možné použiť niekoľko variant oneskorenia, niektorá endogénna premenná môže byť použitá ako exogénna.

V osemdesiatych rokoch navrhol C.A. Sims alternatívny prístup vychádzajúci z vektorovej autoregresie. Dôvodom boli reštrikcie používané pri identifikácii parametrov tradičného modelu, pre ktoré teória podľa Simsa nemá dostatočné oprávnenie.

Vektorová autoregresia je neutrálna voči akejkoľvek špecifickej ekonomickej teórii. Nerozlišuje medzi endogénnymi a exogénnymi premennými, všetky premenné sú endogénne a pracuje sa s nimi symetricky. Známa maximálna dĺžka oneskorenia je rovnaká pre všetky premenné. VAR (vector autoregression) modely sú vhodné na prognózovanie a takisto sa využívajú aj pri analyzovaní hospodárskej politiky, kde sa často skúmajú efekty ekonomických šokov.

3.1 VAR modely – Štrukturálny a redukovaný tvar

Ako sme už spomenuli, viacrovnicový model, v ktorom sú všetky premenné endogénne a každá z nich je vysvetľovaná zvyšnými premennými z toho istého obdobia a o 1,2 až p období oneskorenými hodnotami všetkých premenných modelu sa nazýva vektotrovo autoregresný model rádu p a označuje sa $VAR(p)$:

$$\Gamma y_t = B_0 + B_1 y_{t-1} + B_2 y_{t-2} + \dots + B_p y_{t-p} + u_t, \quad u_t \sim N(0, \Sigma_u) \quad (3.1)$$

Vektor y_t označuje vektor endogénnych premenných v čase t , matica Γ označuje maticu parametrov pri endogénnych premenných, matica B_0 označuje maticu deterministických členov, matica B_1 až B_p označujú matice pri vektoroch oneskorených endogénnych premenných a vektor u_t označuje vektor náhodných zložiek. Náhodé zložky majú normálne rozdelenie s variančno-kovariančnou maticou Σ_u . Tento model predstavuje štrukturálny tvar modelu, ktorý sa nazýva pôvodným systémom. Uvedieme si príklad všeobecného VAR(1) modelu dvoch premenných:

$$\begin{aligned} y_{t,1} &= \beta_{10} + \gamma_{12} y_{t,2} + \beta_{11}^{(1)} y_{t-1,1} + \beta_{12}^{(1)} y_{t-1,2} + \mu_{t,1} \\ y_{t,2} &= \beta_{20} + \gamma_{21} y_{t,1} + \beta_{21}^{(1)} y_{t-1,1} + \beta_{22}^{(1)} y_{t-1,2} + \mu_{t,2} \end{aligned}$$

V tomto modeli predpokladáme, že endogénne premenné sú stacionárne, spĺňajú všetky predpoklady pre náhodné zložky a vzájomne sú nekorelované. Teraz využijeme maticový zápis a príklad v pôvodnom aj maticovom zápise upravíme :

$$\begin{bmatrix} 1 & -\gamma_{12} \\ -\gamma_{21} & 1 \end{bmatrix} \begin{bmatrix} y_{t,1} \\ y_{t,2} \end{bmatrix} = \begin{bmatrix} \beta_{10} \\ \beta_{20} \end{bmatrix} + \begin{bmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \end{bmatrix} \begin{bmatrix} y_{t-1,1} \\ y_{t-1,2} \end{bmatrix} + \begin{bmatrix} \mu_{t,1} \\ \mu_{t,2} \end{bmatrix}$$

$$\Gamma y_t = B_0 + B_1 y_{t-1} + u_t$$

$$\begin{aligned} \begin{bmatrix} y_{t,1} \\ y_{t,2} \end{bmatrix} &= \begin{bmatrix} 1 & -\gamma_{12} \\ -\gamma_{21} & 1 \end{bmatrix}^{-1} \begin{bmatrix} \beta_{10} \\ \beta_{20} \end{bmatrix} + \begin{bmatrix} 1 & -\gamma_{12} \\ -\gamma_{21} & 1 \end{bmatrix}^{-1} \begin{bmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \end{bmatrix} \begin{bmatrix} y_{t-1,1} \\ y_{t-1,2} \end{bmatrix} + \\ &+ \begin{bmatrix} 1 & -\gamma_{12} \\ -\gamma_{21} & 1 \end{bmatrix}^{-1} \begin{bmatrix} \mu_{t,1} \\ \mu_{t,2} \end{bmatrix} \end{aligned}$$

$$y_t = \Gamma^{-1} B_0 + \Gamma^{-1} B_1 y_{t-1} + \Gamma^{-1} u_t$$

$$\begin{aligned} \begin{bmatrix} y_{t,1} \\ y_{t,2} \end{bmatrix} &= \begin{bmatrix} \frac{\beta_{10} + \beta_{20} \gamma_{21}}{1 - \gamma_{12} \gamma_{21}} \\ \frac{\gamma_{12} \beta_{10} + \beta_{20}}{1 - \gamma_{12} \gamma_{21}} \end{bmatrix} + \begin{bmatrix} \frac{\beta_{11} + \gamma_{21} \beta_{21}}{1 - \gamma_{12} \gamma_{21}} & \frac{\beta_{12} + \gamma_{21} \beta_{22}}{1 - \gamma_{12} \gamma_{21}} \\ \frac{\gamma_{12} \beta_{11} + \beta_{21}}{1 - \gamma_{12} \gamma_{21}} & \frac{\gamma_{12} \beta_{12} + \beta_{22}}{1 - \gamma_{12} \gamma_{21}} \end{bmatrix} \begin{bmatrix} y_{t-1,1} \\ y_{t-1,2} \end{bmatrix} + \\ &\begin{bmatrix} \frac{\mu_{t,1} + \mu_{t,2} \gamma_{21}}{1 - \gamma_{12} \gamma_{21}} \\ \frac{\gamma_{21} \mu_{t,1} + \mu_{t,2}}{1 - \gamma_{12} \gamma_{21}} \end{bmatrix} \end{aligned}$$

$$y_t = \Pi_0 + \Pi_1 y_{t-1} + v_t$$

Zo štruktúrného tvaru modelu sme úpravami dostali redukovaný tvar modelu, ktorý sa nazýva štandardným tvarom vektorovo autoregresného modelu:

$$y_{t,1} = \Pi_{10} + \Pi_{11}^{(1)} y_{t-1,1} + \Pi_{12}^{(1)} y_{t-1,2} + v_{t,1}$$

$$y_{t,2} = \Pi_{20} + \Pi_{21}^{(1)} y_{t-1,1} + \Pi_{22}^{(1)} y_{t-1,2} + v_{t,2}$$

kde matice Π_0 a Π_1 označujú matice parametrov redukovaného tvaru modelu a vektor v_t označuje vektor náhodných zložiek tohto modelu.

Pôvodný systém možme analogicky previesť na všeobecný vektorovo autoregresný model rádu p v štandardnom tvare :

$$y_t = \Pi_0 + \Pi_1 y_{t-1} + \Pi_2 y_{t-2} + \dots + \Pi_p y_{t-p} + v_t, \quad v_t \sim N(0, \Sigma_v). \quad (3.2)$$

Kovariančná matica náhodných zložiek má tvar

$$\Sigma_v = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \dots & \sigma_{1n} \\ \sigma_{21} & \sigma_2^2 & \dots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \dots & \sigma_n^2 \end{bmatrix}$$

kde sa dá každý prvok vypočítať ako

$$\sigma_{ij} = \frac{1}{T} \sum_{t=1}^T v_{it} v_{jt}$$

Parametre redukovaného tvaru nemajú iadnu ekonomickú interpretáciu, sú len funkciou parametrov štruktúrného tvaru. Rovnako to platí aj pre náhodné zložky redukovaného tvaru a štruktúrného tvaru, ktoré nazývame štruktúrne šoky.

Napríklad vo $VAR(1)$ modeli s dvoma premennými majú oba štruktúrne šoky vplyv na obe endogénne premenné a vzťah medzi náhodnými zložkami vyzerá nasledovne:

$$\begin{bmatrix} v_{t,1} \\ v_{t,2} \end{bmatrix} \begin{bmatrix} 1 & -\gamma_{12} \\ -\gamma_{21} & 1 \end{bmatrix}^{-1} \begin{bmatrix} u_{t,1} \\ u_{t,2} \end{bmatrix} = \begin{bmatrix} \frac{u_{t,1} + u_{t,2} \gamma_{21}}{1 - \gamma_{12} \gamma_{21}} \\ \frac{\gamma_{21} u_{t,1} + u_{t,2}}{1 - \gamma_{12} \gamma_{21}} \end{bmatrix}$$

Všeobecne sa dá vzťah náhodných zložiek redukovaného a štrukturálneho tvaru dá zapísať nasledovne:

$$v_t = \Gamma^{-1}u_t. \quad (3.3)$$

Procesy v_1 a v_2 sú stacionárne s nulovým priemerom, konštantným rozptylom a nulovými autokoreláciami. Kovariancia medzi nimi však nie je všeobecne nulová a šoky redukovaného tvaru sú korelované :

$$\text{cov}(v_1, v_2) = \frac{(Y_{21}\text{var}(u_1) + Y_{12}\text{var}(u_2))}{(1 - Y_{12}Y_{21})^2}$$

V takom prípade, že by medzi endogénnymi premennými neexistoval žiaden súčasný vplyv (matica Γ by bola diagonálna), šoky redukovaného tvaru nebudú korelované. Pre model s dvomi premennými by platilo $Y_{12} = Y_{21} = 0$. V každej rovnici modelu sú vysvetľujúcimi premennými rovnaké oneskorené endogénne premenné. Keďže o náhodných zložkách predpokladáme, že nie sú autokorelované a rozptyl majú konštantný, najvodnejšou metódou na odhad parametrov je metóda najmenších štvorcov.

3.2 Stabilita

Všeobecne je dynamický systém stabilný vtedy, ak z ľubovoľného počiatočného stavu koverguje k nejakému pevnému bodu. Podmienkou stability v autoregresnom modeli prvého rádu $y_t = a_0 + a_1y_{t-1} + \varepsilon_t$ je, že absolútna hodnota a_1 musí byť menšia než jedna. Ak $|a_1| \leq 1$ bude trajektória časového radu konvergovať, pokiaľ je a_1 kladné, bude mať trajektória tvar exponenciály. Ak by bolo $a_1 > 1$, trajektória bude divergovať a ak by nastal prípad, že $a_1 < -1$, trajektória bude stále viac oscilovať.

Prepisom modelu (3.2) do operátorového tvaru (platí $y_{t-1} = Ly_t$) dostávame tvar

$$y_t = \Pi_0 + \Pi_1Ly_t + \Pi_2L^2y_t + \dots + \Pi_pL^py_t + v_t = \Pi_0 + \sum_{j=1}^p \Pi_jL^jy_t + v_t.$$

Ak využijeme matice operátorového polynómu , dostávame tvar

$$\Pi(L)y_t = \Pi_0 + v_t$$

kde

$$\Pi(L) = I - \Pi_1 L + \Pi_2 L^2 - \dots - \Pi_p L^p$$

Vektorovo autoregresný model rádu p je stabilný, ak korene rovnice

$$|I - \Pi_1 L + \Pi_2 L^2 - \dots - \Pi_p L^p| = 0 \quad (3.4)$$

ležia mimo jednotkového kruhu. Stabilný VAR proces je zároveň stacionárny.

3.3 Woldova dekompozícia

Väčšina ekonomických časových radov je nestacionárna, spravidla vykazujú rastúci trend. Stacionarita sa dá zabezčiť napríklad diferencovaním dát. Niektorí autori však neodporúčajú dáta diferencovať. Argumentujú to tým, že hlavným cieľom VAR analýzy nie je odhad parametrov, ale určenie vzájomných vzťahov medzi premennými. Pri odstraňovaní trendu sa stratí časť informácií o vzájomnom ovplyvňovaní časových radov. Každý stacionárny VAR proces sa dá pomocou Woldovej dekompozície zapísať v tvare kľzavých priemerov nekonečného rádu :

$$y_t = \Pi(L)^{-1}\Pi_0 + \Pi(L)^{-1}v_t = \mu + \sum_{k=0}^{\infty} \theta_k v_{t-k},$$

pričom

$$\theta_0 = I, \quad \lim_{k \rightarrow \infty} \theta_k = 0 \quad (3.5)$$

3.4 Impulzné odozvy (Impulse response)

Woldove koeficienty získame riešením $\Pi(L)\theta(L) = I$, z čoho vyplýva

$$\theta_1 = \Pi_1, \theta_2 = \Pi_1 \Pi_1 + \Pi_2 \text{ až } \theta_k = \Pi_1 \theta_{k-1} + \dots + \Pi_p \theta_{k-p}.$$

Závislosť od štrukturálnych šokov – reprezentáciu SMA získame dosadením (3.3) do (3.5):

$$y_t = \Pi(L)^{-1}\Pi_0 + \Pi(L)^{-1}\Gamma^{-1}u_t = \mu + \sum_{k=0}^{\infty} \Psi_k u_{t-k},$$

$$\text{kde } \Psi_0 = \Gamma^{-1} \text{ a } \Psi_k = \Theta_k \Gamma^{-1} \quad (3.6)$$

Model s dvomi premennými bude mať reprezentáciu kĺzavých priemerov :

$$\begin{bmatrix} y_{t,1} \\ y_{t,2} \end{bmatrix} = \begin{bmatrix} \mu_{10} \\ \mu_{20} \end{bmatrix} + \sum_{k=0}^{\infty} \begin{bmatrix} \theta_{11} & \theta_{12} \\ \theta_{21} & \theta_{22} \end{bmatrix}^{(k)} \begin{bmatrix} v_{t-k,1} \\ v_{t-k,2} \end{bmatrix}.$$

Reprezentácia v závislosti od štrukturálnych šokov nadobúda tvar :

$$\begin{bmatrix} y_{t,1} \\ y_{t,2} \end{bmatrix} = \begin{bmatrix} \mu_{10} \\ \mu_{20} \end{bmatrix} + \sum_{k=0}^{\infty} \begin{bmatrix} \psi_{11}^{(k)} & \psi_{12}^{(k)} \\ \psi_{21}^{(k)} & \psi_{22}^{(k)} \end{bmatrix}^{(k)} \begin{bmatrix} u_{t-k,1} \\ u_{t-k,2} \end{bmatrix}.$$

Prvky matice Ψ_k , teda $\psi_{ij}^{(k)}$ interpretujú reakciu i -tej premennej na jednotkovú zmenu j -teho štrukturálneho šoku po k obdobiach. Všeobecne získame zobrazením $\psi_{ij}^{(k)}$ s meniacim sa k ako krátkodobým multiplikátorom pre jednotlivé obdobia funkciu reakcie i -tej premennej na impulz z j -tej rovnice. Spočítaním reakcií pre všetky k dostávame dlhodobý multiplikátor.

3.5 Grangerova kauzalita

VAR modely sa najčastejšie využívajú pri prognózovaní, kde je dôležité, ktoré premenné zlepšia schopnosť predikcie. Pri rozhodovaní o vhodných premenných sa využíva testovanie kauzality. Overuje sa pri ňom, či oneskorené hodnoty jednej premennej vylepšia predikciu inej premennej. Teraz zapíšeme *VAR* model v upravenom tvare :

$$y_t = \Pi_0 + (1 - \Pi(L))y_t + v_t,$$

kde

$$1 - \Pi(L) = \Pi_1 L + \Pi_2 L^2 + \dots + \Pi_p L^p,$$

čo v našom prípade znamená

$$\begin{bmatrix} y_{t,1} \\ y_{t,2} \end{bmatrix} = \begin{bmatrix} \pi_{10} \\ \pi_{20} \end{bmatrix} + \begin{bmatrix} 1 - \pi_{11}(L) & -\pi_{12}(L) \\ -\pi_{21}(L) & 1 - \pi_{22}(L) \end{bmatrix} \begin{bmatrix} y_{t,1} \\ y_{t,2} \end{bmatrix} + \begin{bmatrix} v_{t,1} \\ v_{t,2} \end{bmatrix}.$$

Ak berieme do úvahy model s dvomi premennými, tak Y_1 nevplyva na Y_2 v Grangerovom zmysle ak sa všetky koeficienty $-\pi_{21}$ rovnajú nule, čo testujeme F-štatistikou pomocou hypotézy:

$$-\pi_{21}(1) = -\pi_{21}(2) = \dots = -\pi_{21}(p) = 0.$$

Vo všeobecnom modeli testujeme, či premenná Y_i nevlýva na ostatné premenné, čo platí práve vtedy, ak sa všetky koeficienty i -tej premennej rovnajú nule. Teda $-\pi_{ji} = 0$ pre všetky j .

Kapitola 4

Model lineárnej regresie

4.1 Identifikácia modelu

Jednou z najťažších úloh pri tvorbe Box-Jenkinsových modelov je ich identifikácia. Táto úloha spočíva v rozhodnutí, aký typ modelu vybrať. Je to však iba prvá fáza konštrukcie modelu. Identifikáciu modelu $ARIMA(p, d, q)$ si vieme zhrnúť do nasledujúcich krokov:

1. Preskúmanie grafu časového radu. V mnohých prípadoch je možné na prvý pohľad rozpoznať napríklad trend, alebo sezónnosť v časovom rade. V tejto fáze ide predovšetkým o subjektívne zhodnotenie situácie. Je vhodné previesť rôzne úpravy, napríklad previesť transformáciu, ktorá stabilizuje rozptyl.
2. Ďalším krokom je výpočet ACF a PACF. Na ich základe môžeme prepokladať, či je časový rad stacionárny, alebo nestacionárny. Charakteristickým znakom nestacionarity je pomalé klesanie ACF.
3. Prípadá stacionarita sa dá odstrániť napríklad diferencovaním. Po stacionarizácii radu sa pre určenie parametrov p, q odhadne ACF a PACF, ktoré sa porovnávajú s teoretickými ACF a PACF.

Pre konštrukciu vierohodného modelu sa odporúča použiť aspoň $T = 50$ pozorovaní. Odhad ACF a PACF by malo tvoriť asi $T/4$ hodnôt.

4.2 Odhad parametrov

Metóda najmenších štvorcov

V ekonometrii patrí táto metóda odhadu parametrov k najčastejším. Spočíva v minimalizácii súčtu štvorcov odchýlok, ktoré sú určené ako rozdiel medzi pozorovanými a vypočítanými hodnotami vysvetľovanej premennej. Uvedieme si príklad na stochastickej špecifikácii jednoduchého lineárneho modelu:

$$y_i = \beta_0 + \beta_1 x_i + u_i$$

Odhadnutá regresná funkcia bude mať tvar

$$y_i = \hat{\beta}_0 + \hat{\beta}_1 x_i + \hat{u}_i = \hat{y}_i + \hat{u}_i$$

Pretože náhodná zložka sa priamo nikdy neodhaduje, ale jej odhad sa vypočíta ako rozdiel hodnôt y_i a \hat{y}_i , namiesto označenia \hat{u}_i budeme používať e_i a tento rozdiel nazveme reziduál. Následne aj stochastická špecifikácia regresnej funkcie bude mať tvar:

$$y_i = \hat{\beta}_0 + \hat{\beta}_1 x_i + e_i = \hat{y}_i + e_i,$$

teda

$$e_i = y_i - \hat{y}_i = y_i - \hat{\beta}_0 + \hat{\beta}_1 x_i$$

Minimalizáciu štvorcov reziduálov teda môžeme zapísať nasledovne:

$$\begin{aligned} \min_{\hat{\beta}_0, \hat{\beta}_1} \sum_{i=1}^n e_i^2 &= \min_{\hat{\beta}_0, \hat{\beta}_1} \sum_{i=1}^n (y_i - \hat{\beta}_0 + \hat{\beta}_1 x_i)^2 = \\ &= \min_{\hat{\beta}_0, \hat{\beta}_1} \left(\sum_{i=1}^n y_i^2 - 2\hat{\beta}_0 \sum_{i=1}^n y_i + n\hat{\beta}_0^2 - 2\hat{\beta}_1 \sum_{i=1}^n x_i y_i + 2\hat{\beta}_0 \hat{\beta}_1 \sum_{i=1}^n x_i + \hat{\beta}_1^2 \sum_{i=1}^n x_i^2 \right). \end{aligned}$$

Výhodou tejto metódy je, že s rastom odchýlenia zvyšuje minimalizovanú hodnotu kvadraticky, teda v neprospech vyšších odchýlok oproti relatívnej rovnocennosti zvýšenia tejto hodnoty v prípade minimalizovania súčtu absolútnych hodnôt reziduálov.

4.3 Verifikácia modelu

4.3.1 Stacionarita – Dickey-Fullerov test

Väčšina časových radov v praxi je nestacionárna. Môže sa jednať o deterministickú stacionaritu, ktorú spôsobuje trend, alebo o stochastickú stacionaritu, ktorú môžeme vyriešiť diferencovaním časového radu. Tento spôsob stacionarizovania svedčí o prítomnosti jednotkového koreňa. Najpoužívanejším testom na overenie prítomnosti jednotkového koreňa je Dickey-Fullerov test.

Uvažujme proces AR(1)

$$x_t = a_1 x_{t-1} + \varepsilon_t \quad (4.1)$$

Ak $a_1 = 1$, rovnica obsahuje jednotkový koreň, čiže má tvar

$$x_t = x_{t-1} + \varepsilon_t.$$

V základnom teste na testovanie prítomnosti jednotkového koreňa testujeme hypotézu:

H_0 : časový rad obsahuje jednotkový koreň ($a_1 = 1$)

proti alternatíve

H_1 : časový rad neobsahuje jednotkový koreň ($a_1 < 1$)

Ak v rovnici (4.1) odčítame z oboch strán x_{t-1} dostávame

$$\Delta x_t = \delta x_{t-1} + \varepsilon_t, \quad (4.2)$$

teda sa nám hypotézy zmenia nasledovne:

H_0 : časový rad obsahuje jednotkový koreň ($\delta = 0$) (4.3)

proti alternatíve

H_1 : časový rad neobsahuje jednotkový koreň ($\delta < 0$)

Metódou najmenších štvorcov odhadneme rovnicu(4.2) a určíme testovaciu t-štatistiku pre parameter δ ako

$$t_\delta = \frac{\hat{\delta}}{se(\hat{\delta})}, \quad (4.4)$$

pričom $\hat{\delta}$ je odhad pre δ a $se(\hat{\delta})$ je štandardná odchýlka tohto odhadu.

Dickey a Fuller zistili, že za platnosti nulovej hypotézy (4.3) táto t-štatistika nekopíruje štandardné Studentovo rozdelenie. Najprv pre túto štatistiku odvodili asymptotické rozdelenie a neskôr pomocou väčšieho počtu simulácií MacKinnon spresnil jej kritické hodnoty. Nevýhodou klasického Dickey-Fullerovho testu je využitie len pre AR(1) procesy. Preto Dickey a Fuller vyvinuli rozšírený test (ADF), kde rozšírili odhadovanú rovniu pre AR(p) procesy, pričom za nulovú hypotézu budeme pokladať:

$$H_0: \Delta x_t = \delta x_{t-1} + \sum_{i=1}^p \gamma_i \Delta x_{t-1} + \varepsilon_t, \quad \text{pre } \delta = 0.$$

4.3.2 Koeficient determinácie

Koeficient determinácie označovaný ako R^2 nám vyjadruje podiel celkovej variability závislej premennej ktorú vysvetľuje model. Jeho hodnota je vyjadrená v percentách, je v rozmedzí od 0 po 1 pričom vyššia hodnota signalizuje lepší model. Počíta podľa vzťahu:

$$R^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS}$$

pričom RSS (residual sum of squares) je reziduálny súčet štvorcov a počíta sa nasledovne:

$$RSS = \sum_{t=1}^T \hat{\epsilon}_t^2 = \sum_{t=1}^T (y_t - \hat{y}_t)^2$$

Čím je jeho nezáporná hodnota menšia, tým je model lepší. Úplný súčet štvorcov (total sum of squares) počítame podľa vzťahu:

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

Vysvetlený súčet štvorcov (explained sum of squares) dostávame zo vzťahu:

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

Platí

$$TSS = ESS + RSS.$$

V praxi sa používa aj korigovaný koeficient determinácie:

$$\bar{R}^2 = 1 - \left[\frac{T-1}{T-k} (1 - R^2) \right]$$

pričom k je počet vysvetľujúcich premenných.

4.3.3 Heteroskedasticita

Jednou zo základných požiadaviek lineárneho regresného modelu je homoskedasticita, čo znamená že náhodná zložka, teda aj reziduá modelu sú konečné a konštantné. Porušenie tohto predpokladu sa nazýva heteroskedasticita., ktorú budeme testovať White heteroskedasticity testom.

4.3.4 Autokorelovanosť reziduí

Autokorelovanosť náhodných zložiek môžeme modelovať nasledovne :

$$\epsilon_t = \rho \epsilon_{t-1} + \mu_t$$

pričom parameter ρ označíme ako biely šum. Takýto model nazývame autorehresným modelom prvého rádu. Parameter ρ nadobúda hodnoty od -1 po 1. Hodnota 0 signalizuje nekorelovanosť a hodnoty priplížujúce sa k -1 resp. 1 signalizujú zápornú resp. Kladnú koreláciu. Na testovanie autokorelácie budeme používať Durbin-Watsonov test. Nulová hypotéza bude:

$$H_0: \rho = 0$$

Testovacia štatistika:

$$DW = \frac{\sum_{t=2}^T (\hat{\epsilon}_t - \hat{\epsilon}_{t-1})^2}{\sum_{t=1}^T \hat{\epsilon}_t^2}$$

Vysoké hodnoty znamenajú zápornú koreláciu, naopak nízke hodnoty predstavujú kladnú koreláciu. Aproximáciou dostávame:

$$DW = 2(1 - \hat{\rho}),$$

pričom

$$\hat{\rho} = \frac{\sum_{t=2}^T \hat{\epsilon}_t \hat{\epsilon}_{t-1}}{\sum_{t=1}^T \hat{\epsilon}_t^2}.$$

Teda platí

- $DW = 0$, ak $\rho = 1$
- $DW = 2$, ak $\rho = 0$
- $DW = 4$, ak $\rho = -1$

4.3.5 Informačné kritériá pre výber modelu

Počas tvorby modelu sa strtávame so situáciou, kedy sa rozhodujeme či pridať do modelu ešte nejakú vysvetľujúcu premennú. Dôsledkom takéhoto rozšírenia modelu môže byť pokles súčtu štvorcov reziduí, no riskujeme pri tom preparametrizovanie modelu, inými slovami použitia viac vysvetľujúcich premenných v modeli, ako by bolo potrebné. Takýmto spôsobom sa totiž náš model môže vzdialiť od toho správneho.

Informačné kritériá nám slúžia práve na takéto porovnanie modelov. Predstavíme si najpoužívanejšie informačné kritériá:

Akaikeho informačné kritérium: $AIC = 2k + n[\ln(RSS/n)]$

Schwarzovo informačné kritérium: $SIC = k \ln(n) + n[\ln(RSS/n)]$

Hannan-Quinnovo informačné kritérium: $HQC = 2k \ln(\ln(n)) + n[\ln(RSS/n)]$

Pričom k je počet parametrov a n je počet pozorovaní.

Porovnávané modely musia byť odhadnuté na rovnakých dátach a premenné musia byť v rovnakom tvare. Snažíme sa minimalizovať informačné kritériá. V prípade, že pridaná premenná nemá dostatočnú vysvetľujúcu silu, hodnota informačného kritéria sa zvýši. Takýmto spôsobom sa vlastne penalizuje zavedenie zbytočného parametra do modelu.

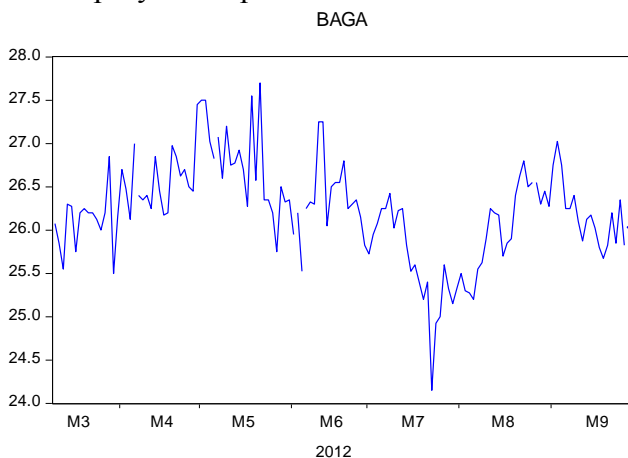
Kapitola 5

Praktická časť

V tejto kapitole sa budeme snažiť vytvoriť modely, pomocou ktorých dokážeme prognózovať cenu zemného plynu tak, aby sa čo najviac priblížila pozorovanej cene. K dispozícii máme 144 pozorovaní, ktoré vznikli počas obchodných dní v období 1.3.2012-28.9.2012. Ceny zemného plynu sú z obchodných miest so zemným plynom (HUBov) v Rakúsku (BAGA) , Nemecku (Gaspool, NCG), Holandsku (TTF) a Veľkej Británii (NBP). V prvej časti vytvoríme ARIMA modely pre trhy BAGA a NCG, do ktorých potom zahrnieme exogénne premenné a budeme skúmať ich vplyv na model. V ďalšej časti sa budeme zaoberať VAR modelom, ktorý bude zahŕňať všetky tieto pozorovania ako endogénne premenné. Na záver porovnáme tieto dva prístupy. Budeme pracovať v programe Ewiews v.7. Korelogramy ACF a PACF nebudeme zobrazovať v celej dĺžke z dôvodu nesignifikantnosti vo vyšších rádoch.

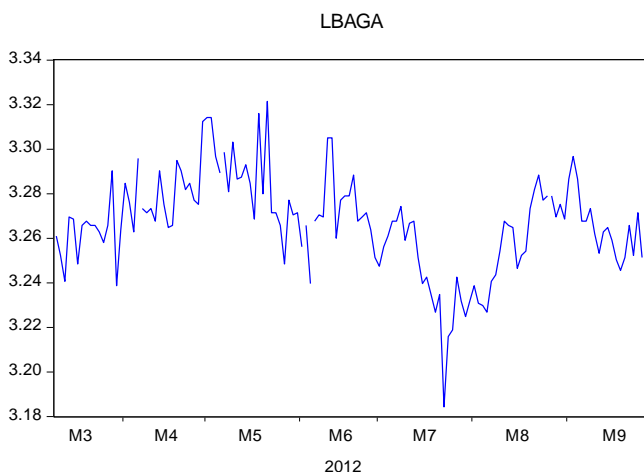
5.1 ARIMA modelovanie

Ako prvý sa pokúsime namodelovať hub BAGA. Vykreslíme si jeho graf.



obr.5.1:ceny BAGA

Pre vyhladenie dát použijeme prirodzené logaritmy:



obr.5.2:lnBAGA

Otestujeme stacionaritu časového radu pomocou ADF testu:

Null Hypothesis: LBAGA has a unit root
 Exogenous: None
 Lag Length: 1 (Automatic - based on SIC, maxlag=13)

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-0.004899	0.6794
Test critical values:		
1% level	-2.581827	
5% level	-1.943157	
10% level	-1.615178	

*Mackinnon (1996) one-sided p-values.

obr.5.3: ADF test lnBAGA

p-value 0,6794 nám nezamieta nulovú hypotézu o prítomnosti jednotkového koreňa, teda časový rad je nestacionárny, čo sme aj očakávali, pretože sa jedná o denné finančné dáta, pre ktoré je charakteristická stochastická nestacionarita. Otestujeme zdiferencovaný rad:

Null Hypothesis: D(LBAGA) has a unit root
 Exogenous: None
 Lag Length: 0 (Automatic - based on SIC, maxlag=13)

















	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-18.47159	0.0000
Test critical values:		
1% level	-2.581827	
5% level	-1.943157	
10% level	-1.615178	

*Mackinnon (1996) one-sided p-values.

obr.5.4: ADF test dlnBAGA

Z tohto výstupu vidno, že časový rad sa zdiferencovaním stal stacionárnym. Budeme preto pracovať s prvými diferenciami logaritmov cien BAGA. Vykreslíme korelogram pre znázornenie ACF a PACF:

Date: 04/23/13 Time: 19:59
 Sample: 3/08/2012 9/28/2012
 Included observations: 139

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob
		1 -0.415	-0.415	24.510	0.000
		2 0.022	-0.182	24.578	0.000
		3 -0.033	-0.122	24.733	0.000
		4 -0.032	-0.120	24.880	0.000
		5 -0.017	-0.114	24.922	0.000
		6 0.078	0.012	25.821	0.000
		7 -0.032	-0.000	25.976	0.001
		8 -0.013	-0.027	26.001	0.001

obr.5.5: korelogram dlnbag

Na prvý pohľad môžeme usúdiť, že sa jedná o ARIMA(2,1,1) proces, no pri tvorbe modelu sme sa riadili hlavne maximalizovaním R-squared a minimalizovaním SIC. Pri takomto prístupe sa ako najlepší model javí ARIMA(3,1,3). Výstup odhadnutej rovnice je nasledovný:

Dependent Variable: D(LBAGA)
 Method: Least Squares
 Date: 04/23/13 Time: 20:06
 Sample: 3/07/2012 9/28/2012
 Included observations: 140
 Convergence achieved after 28 iterations
 MA Backcast: 3/02/2012 3/06/2012

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.000110	0.000502	0.219191	0.8268
AR(1)	-0.419332	0.157668	-2.659597	0.0088
AR(2)	-0.681941	0.116080	-5.874758	0.0000
AR(3)	0.168262	0.144760	1.162352	0.2472
MA(1)	-0.108003	0.121967	-0.885509	0.3775
MA(2)	0.585253	0.070206	8.336175	0.0000
MA(3)	-0.662737	0.116510	-5.688249	0.0000
R-squared	0.334985	Mean dependent var		0.000139
Adjusted R-squared	0.304984	S.D. dependent var		0.016754
S.E. of regression	0.013968	Akaike info criterion		-5.655449
Sum squared resid	0.025947	Schwarz criterion		-5.508367
Log likelihood	402.8815	Hannan-Quinn criter.		-5.595680
F-statistic	11.16591	Durbin-Watson stat		1.949284
Prob(F-statistic)	0.000000			
Inverted AR Roots	.21	-.31+.84i	-.31-.84i	
Inverted MA Roots	.68	-.29+.94i	-.29-.94i	

obr.5.6: model pre dlnbaga

Overíme invertovateľnosť a stacionaritu modelu:

Inverse Roots of AR/MA Polynomial(s)
 Specification: D(LBAGA) C AR(1) AR(2) AR(3)
 MA(1) MA(2) MA(3)
 Date: 04/23/13 Time: 20:17
 Sample: 3/07/2012 9/28/2012
 Included observations: 140

AR Root(s)	Modulus	Cycle
-0.313299 ± 0.844782i	0.901007	3.262406
0.207266	0.207266	
No root lies outside the unit circle. ARMA model is stationary.		
MA Root(s)	Modulus	Cycle
-0.286073 ± 0.944754i	0.987116	3.369322
0.680150	0.680150	

No root lies outside the unit circle.
 ARMA model is invertible.

obr.5.7: ARMA structure

Program EViews pracuje s inverznými koreňmi, takže podmienky stacionarity a invertovateľnosti sú splnené, ak tieto inverzné korene ležia vo vnútri jednotkového kruhu.

Z hodnoty *Durbin-Watsonovej* štatistiky 1,949 môžeme usúdiť, že reziduá nie sú autokorelované, teda sú bielym šumom, čo si môžeme potvrdiť aj vykreslením korelogramu reziduí, v ktorom si budeme všímať hodnoty *Ljung-Boxovej Q-štatistiky*. Nulovou hypotézou je, že korelácia do rádu k je nulová. Ak sa hypotéza nezamieta, model je v poriadku.

Date: 04/23/13 Time: 20:43
 Sample: 3/07/2012 9/28/2012
 Included observations: 140
 Q-statistic probabilities adjusted for 6 ARMA term(s)

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
		1	0.020	0.020	0.0555	
		2	-0.004	-0.004	0.0573	
		3	-0.037	-0.037	0.2585	
		4	0.036	0.037	0.4447	
		5	0.009	0.008	0.4579	
		6	0.037	0.036	0.6624	
		7	0.044	0.046	0.9568	0.328
		8	0.062	0.060	1.5373	0.464
		9	0.054	0.054	1.9728	0.578
		10	-0.116	-0.118	4.0209	0.403
		11	-0.052	-0.048	4.4404	0.488
		12	-0.074	-0.078	5.2841	0.508
		13	0.055	0.041	5.7591	0.568
		14	-0.040	-0.045	6.0094	0.646
		15	0.122	0.121	8.3874	0.496
		16	-0.006	0.000	8.3933	0.590
		17	0.031	0.039	8.5509	0.663
		18	-0.092	-0.068	9.9313	0.622
		19	0.027	0.046	10.052	0.690
		20	0.022	0.015	10.129	0.753
		21	-0.130	-0.159	12.932	0.608
		22	-0.091	-0.117	14.342	0.573
		23	-0.043	-0.056	14.657	0.620
		24	-0.089	-0.130	16.017	0.591
		25	-0.000	0.039	16.017	0.656
		26	-0.066	-0.058	16.772	0.668
		27	-0.111	-0.065	18.950	0.588
		28	0.060	0.065	19.584	0.609

obr.5.8:korelogram rezidui

Z výstupu vidno, že model je vyhovujúci. Teraz overíme heteroskedasticitu. Použijeme *White heteroskedasticity test*, v ktorom testujeme nulovú hypotézu o neprítomnosti heteroskedasticity. Zaujímá nás p-hodnota v druhom riadku.

Heteroskedasticity Test: White

F-statistic	0.978684	Prob. F(35,104)	0.5126
Obs*R-squared	34.68656	Prob. Chi-Square(35)	0.4831
Scaled explained SS	39.50457	Prob. Chi-Square(35)	0.2756

obr.5.9: White test

Z tohto výstupu vidno, že v modeli sa nenachádza heteroskedasticita, čo je pozitívne.

Teraz náš model rozšírime pridaním exogénnej premennej, pričom sa samozrejme budeme snažiť o jeho vylepšenie a zároveň dodržanie všetkých požiadaviek pre vyhovujúci model.

Postupným skúšaním pridávania premenných, ktoré musia byť takisto stacionárne, sme dospeli k nasledujúcemu modelu v ktorom pribudla premenná DTF(-1), ktorá reprezentuje zdiferencované logaritmy cien na TTF posunuté o jedno pozorovanie do minulosti:

Dependent Variable: D(LBAGA)
 Method: Least Squares
 Date: 04/23/13 Time: 21:08
 Sample (adjusted): 3/08/2012 9/28/2012
 Included observations: 139 after adjustments
 Convergence achieved after 14 iterations
 MA Backcast: 3/05/2012 3/07/2012

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-4.48E-05	0.000475	-0.094273	0.9250
DTTF(-1)	0.192932	0.048366	3.989020	0.0001
AR(1)	-0.349521	0.148943	-2.346670	0.0204
AR(2)	-0.665442	0.103306	-6.441438	0.0000
AR(3)	0.091824	0.136696	0.671741	0.5029
MA(1)	-0.199212	0.115347	-1.727067	0.0865
MA(2)	0.650601	0.056567	11.50142	0.0000
MA(3)	-0.659940	0.112293	-5.876972	0.0000
R-squared	0.392243	Mean dependent var		0.000182
Adjusted R-squared	0.359767	S.D. dependent var		0.016807
S.E. of regression	0.013448	Akaike info criterion		-5.724147
Sum squared resid	0.023691	Schwarz criterion		-5.555256
Log likelihood	405.8282	Hannan-Quinn criter.		-5.655514
F-statistic	12.07808	Durbin-Watson stat		1.997592
Prob(F-statistic)	0.000000			
Inverted AR Roots	.13	-.24+.82i	-.24-.82i	
Inverted MA Roots	.68	-.24+.96i	-.24-.96i	

obr.5.10: model pre BAGA

Vidíme, že oproti pôvodnému modelu sa nám zvýšili koeficienty determinácii a zlepšili sa nám aj hodnoty informačných kritérií, čo potvrdzuje celkové zlepšenie modelu.

Teraz overíme či reziduá rozšíreného modelu nie sú korelované a či majú konštantný rozptyl:

Heteroskedasticity Test: White

F-statistic	0.698084	Prob. F(44,94)	0.9070
Obs*R-squared	34.23372	Prob. Chi-Square(44)	0.8548
Scaled explained SS	45.83836	Prob. Chi-Square(44)	0.3958

obr.5.11: White test

Heteroskedasticita nie je prítomná, rozptyl je konštantný.

Hodnota *Durbin-Watsonovej* štatistiky 1,997 zamieta autokoreláciu reziduí. Potvrdíme si to ešte *Q*-štatistikou:

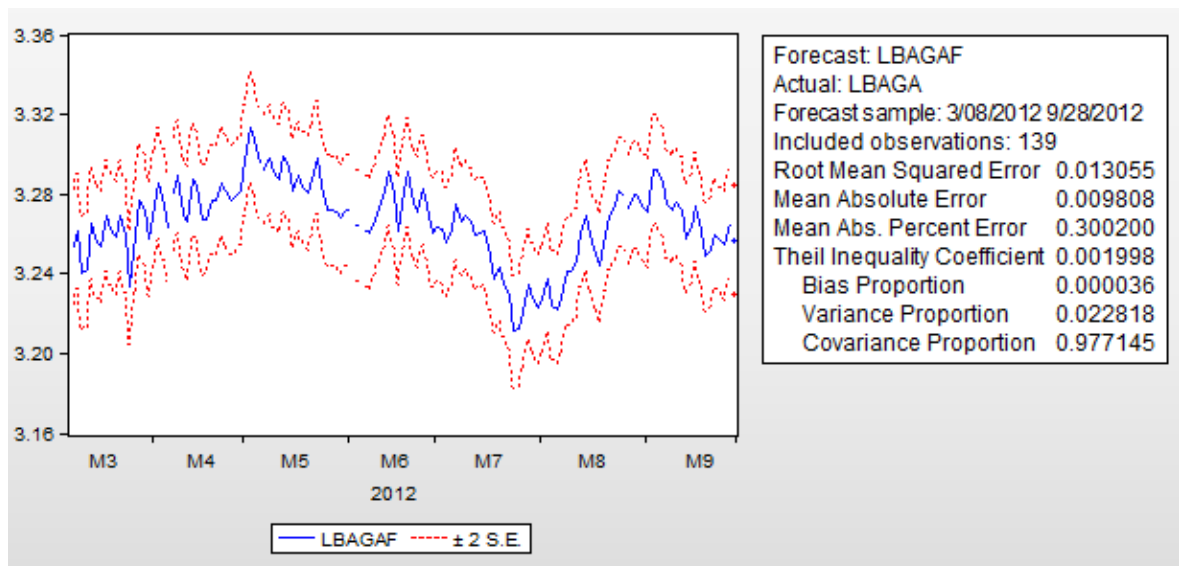
Date: 04/23/13 Time: 21:24
 Sample: 3/08/2012 9/28/2012
 Included observations: 139
 Q-statistic probabilities adjusted for 6 ARMA term(s)

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
		1	-0.001	-0.001	0.0002	
		2	-0.022	-0.022	0.0701	
		3	-0.011	-0.011	0.0871	
		4	0.045	0.045	0.3847	
		5	-0.028	-0.029	0.5005	
		6	0.054	0.056	0.9330	
		7	0.058	0.058	1.4273	0.232
		8	0.049	0.050	1.7924	0.408
		9	0.027	0.034	1.9038	0.593
		10	-0.113	-0.117	3.8558	0.426
		11	-0.058	-0.060	4.3721	0.497
		12	-0.066	-0.078	5.0436	0.538
		13	0.037	0.025	5.2514	0.629
		14	-0.055	-0.057	5.7317	0.677
		15	0.155	0.152	9.5369	0.389
		16	0.069	0.086	10.302	0.414
		17	-0.048	-0.029	10.679	0.471
		18	-0.095	-0.063	12.150	0.434
		19	0.045	0.040	12.477	0.489
		20	0.022	0.017	12.559	0.561
		21	-0.163	-0.195	16.999	0.319
		22	-0.021	-0.065	17.070	0.381
		23	-0.018	-0.053	17.127	0.446
		24	-0.142	-0.178	20.587	0.301
		25	-0.045	-0.002	20.940	0.340
		26	-0.043	-0.018	21.255	0.382
		27	-0.170	-0.141	26.321	0.194
		28	0.047	0.057	26.707	0.223
		29	0.008	0.045	26.719	0.268
		30	0.015	0.044	26.758	0.316
		31	-0.023	-0.038	26.854	0.363
		32	0.089	0.093	28.302	0.344
		33	0.068	0.116	29.162	0.353

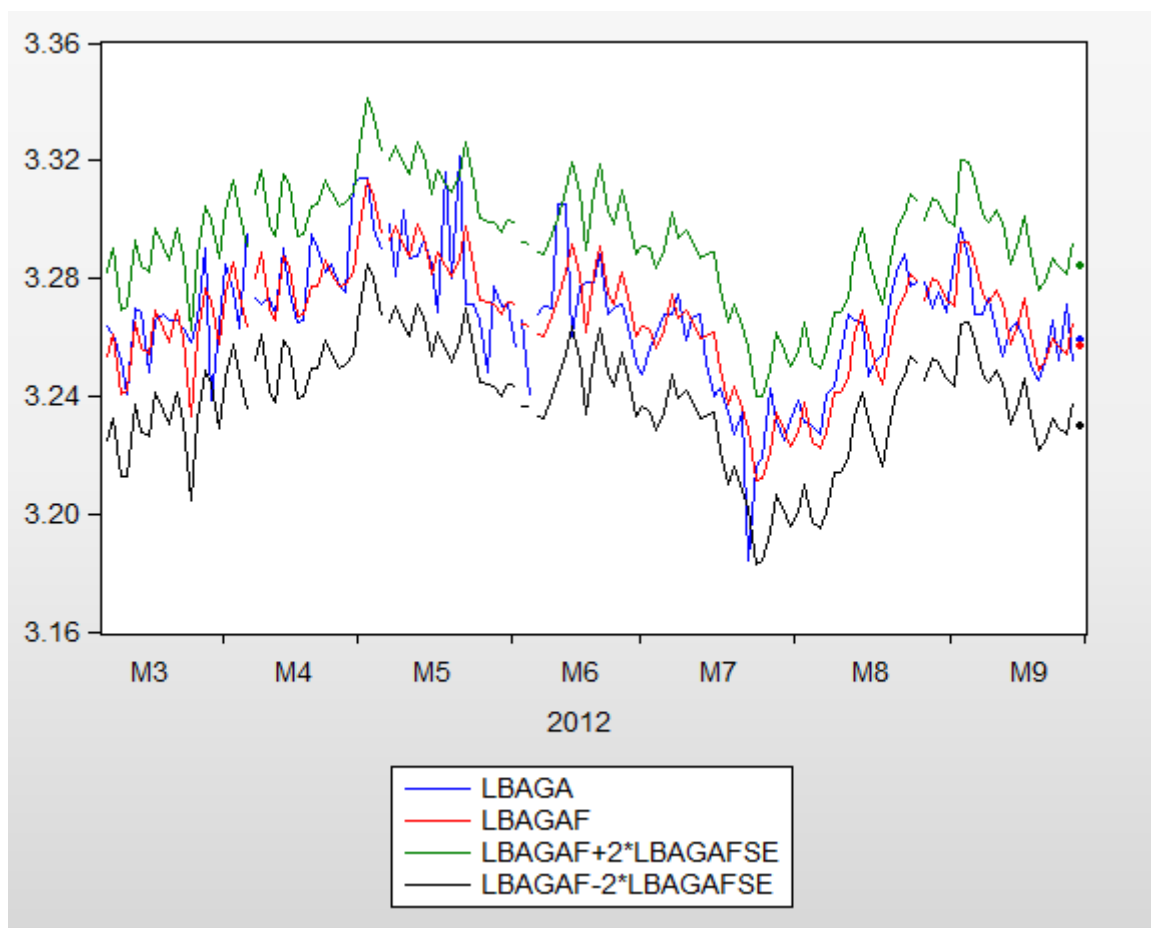
obr.5.12:korelogram rezidui

Zo všetkých výstupov usudzujeme že model je vyhovujúci. Môžeme teraz pristúpiť k predikciám z modelu, ktoré následne porovnáme s pozorovanými hodnotami.

Pri tvorení predikcii budeme používať statické predovede, ktoré každú predpovedanú hodnotu počítajú na základe modelu zo skutočných hodnôt, ktoré máme k dispozícii. Získané predpovede si zobrazíme na grafe spolu so skutočnými pozorovaniami.



obr.5.13: forecast lbaga



obr.5.14: porovnanie predikcie s pozorovaniami BAGA

Vysvetlíme si niektoré pojmy ktoré máme k dispozícii vo výstupe predpovede:

Uvažujeme rozsah predikcie $j = T + 1, T + 2, \dots, T + h$, skutočnú hodnotu pozorovania v čase t ako y_t a predikovanú hodnotu ako \hat{y}_t .

- Root Mean Squared Error = $\sqrt{\sum_{t=T+1}^{T+h} \frac{(\hat{y}_t - y_t)^2}{h}}$.
- Mean Absolute Error = $\sum_{t=T+1}^{T+h} |\hat{y}_t - y_t| / h$.
- Mean Absolute Percentage Error = $100 \sum_{t=T+1}^{T+h} \left| \frac{\hat{y}_t - y_t}{y_t} \right| / h$.

- Theil Inequality Coefficient = $\frac{\sqrt{\sum_{t=T+1}^{T+h} \frac{(\hat{y}_t - y_t)^2}{h}}}{\sqrt{\sum_{t=T+1}^{T+h} \frac{(\hat{y}_t)^2}{h} + \sum_{t=T+1}^{T+h} \frac{(y_t)^2}{h}}}$.

- Bias Proportion = $\frac{((\sum_{t=T+1}^{T+h} \frac{\hat{y}_t}{h}) - \bar{y}_t)^2}{\sum_{t=T+1}^{T+h} (\hat{y}_t - y_t)^2 / h}$.

- Variance Proportion = $\frac{(s_{\hat{y}} - s_y)^2}{\sum_{t=T+1}^{T+h} (\hat{y}_t - y_t)^2 / h}$.

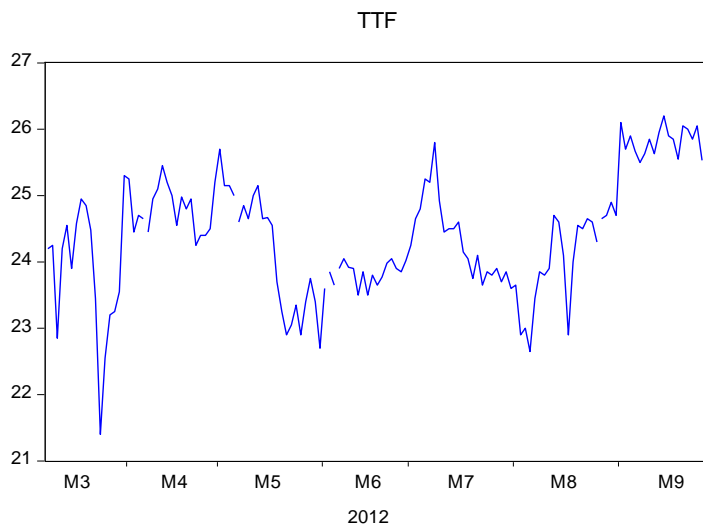
- Covariance Proportion = $\frac{2(1-r)s_{\hat{y}}s_y}{\sum_{t=T+1}^{T+h} (\hat{y}_t - y_t)^2 / h}$.

pričom $\frac{\hat{y}_t}{h}$, y_t , $s_{\hat{y}}$, s_y sú priemery a štandardné odchýlky pre \hat{y}_t a y a r je korelácia medzi \hat{y} a y .

Čím je hodnota RMSE menšia, model lepšie predpovedá do budúcnosti, takisto to platí aj pre hodnotu MSE. Hodnota Theilovho koeficientu leží medzi 0 a 1, čím bližšie je k nule, tým lepšiu zhodu sme dosiahli. Posledné hodnoty dávajú súčet rovný 1. Čím nižšia hodnota Bias, tým lepší model sme odhadli.

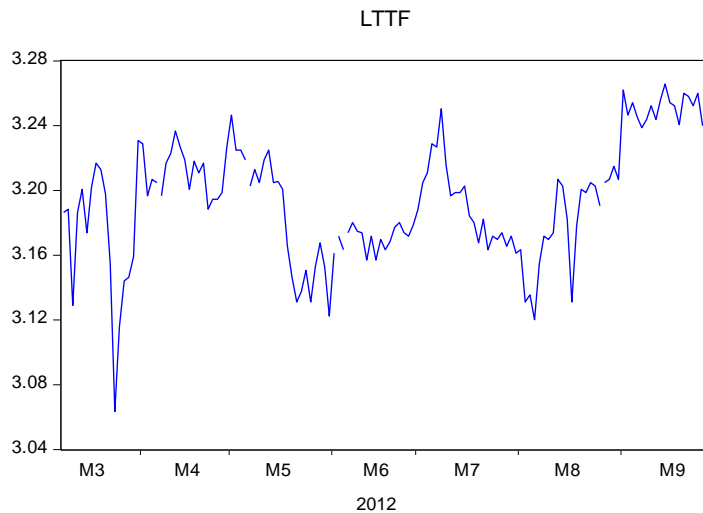
Na základe hodnôt uvedených štatistík predpovede a zhladnutí grafov na ktorých vidno že všetky skutočné pozorovania sa nachádzajú vo vnútri intervalu spoľhlivosti, môžeme konštatovať, že sme skonštruovali dobrý model.

Ako druhý si odhadneme model pre TTF:



obr.5.15:ceny TTF

Po transformácii na prirodzené logaritmy:



obr.5.16:ln TTF

Otestujeme stacionaritu:

Null Hypothesis: LTTF has a unit root
 Exogenous: None
 Lag Length: 0 (Automatic - based on SIC, maxlag=13)

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	0.295487	0.7700
Test critical values:		
1% level	-2.581705	
5% level	-1.943140	
10% level	-1.615189	

*MacKinnon (1996) one-sided p-values.

obr.5.17: ADF lnTTF

Časový rad je nestacionárny, preto ho diferencujeme a znovu vykonáme ADF test:

Null Hypothesis: D(LTTF) has a unit root
 Exogenous: None
 Lag Length: 0 (Automatic - based on SIC, maxlag=13)

















	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-13.00268	0.0000
Test critical values:		
1% level	-2.581705	
5% level	-1.943140	
10% level	-1.615189	

*MacKinnon (1996) one-sided p-values.

obr.5.18: ADF DlnTTF

Zdiferencovaný časový rad je už stacionárny. Zobrazíme ACF a PACF :

Date: 04/23/13 Time: 23:11
 Sample: 3/08/2012 9/28/2012
 Included observations: 139

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob
		1 -0.104	-0.104	1.5406	0.215
		2 -0.161	-0.174	5.2431	0.073
		3 0.020	-0.019	5.2990	0.151
		4 0.017	-0.012	5.3388	0.254
		5 -0.244	-0.254	14.078	0.015
		6 0.004	-0.063	14.080	0.029
		7 -0.051	-0.159	14.468	0.043
		8 0.079	0.034	15.412	0.052

obr.5.19: ACF a PACF

Najlepšie hodnoty R-squared a informačných kritérií sme dosiahli pri modeli ARIMA(3,1,3) :

Dependent Variable: D(LTTF)
 Method: Least Squares
 Date: 04/23/13 Time: 23:18
 Sample: 3/08/2012 9/28/2012
 Included observations: 139
 Convergence achieved after 35 iterations
 MA Backcast: 3/05/2012 3/07/2012

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.000439	0.000248	1.770780	0.0789
AR(1)	-0.024981	0.059042	-0.423103	0.6729
AR(2)	-0.135826	0.056545	-2.402070	0.0177
AR(3)	0.704046	0.055497	12.68610	0.0000
MA(1)	-0.003461	0.019205	-0.180232	0.8572
MA(2)	-0.002552	0.019335	-0.132008	0.8952
MA(3)	-0.975183	0.014297	-68.20876	0.0000
R-squared	0.253363	Mean dependent var		0.000597
Adjusted R-squared	0.219425	S.D. dependent var		0.021138
S.E. of regression	0.018675	Akaike info criterion		-5.074200
Sum squared resid	0.046036	Schwarz criterion		-4.926421
Log likelihood	359.6569	Hannan-Quinn criter.		-5.014147
F-statistic	7.465445	Durbin-Watson stat		2.088012
Prob(F-statistic)	0.000001			
Inverted AR Roots	.83	-.43-.81i	-.43+.81i	
Inverted MA Roots	.99	-.50+.86i	-.50-.86i	

obr.5.20: ARIMA DlnTTF

Najviac sa nám podarilo vylepšiť model rozšírením o premennú DNBP, ktorá predstavuje zdiferencované logaritmy cien na hube NBP :

Dependent Variable: D(LTTF)
 Method: Least Squares
 Date: 04/23/13 Time: 23:29
 Sample: 3/08/2012 9/28/2012
 Included observations: 139
 Convergence achieved after 21 iterations
 MA Backcast: 3/07/2012

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.000162	3.97E-05	4.089110	0.0001
DNBP	0.760089	0.026373	28.82040	0.0000
AR(1)	0.194745	0.085739	2.271380	0.0247
MA(1)	-0.979125	0.011992	-81.64739	0.0000
R-squared	0.729215	Mean dependent var		0.000597
Adjusted R-squared	0.723197	S.D. dependent var		0.021138
S.E. of regression	0.011121	Akaike info criterion		-6.131619
Sum squared resid	0.016696	Schwarz criterion		-6.047173
Log likelihood	430.1475	Hannan-Quinn criter.		-6.097302
F-statistic	121.1833	Durbin-Watson stat		2.002620
Prob(F-statistic)	0.000000			
Inverted AR Roots	.19			
Inverted MA Roots	.98			

obr.5.21: model pre TTF

Overíme stacionaritu a invertovateľnosť modelu:

Inverse Roots of AR/MA Polynomial(s)
 Specification: D(LTTF) C AR(1) MA(1) DNBP
 Date: 04/23/13 Time: 23:34
 Sample: 3/08/2012 9/28/2012
 Included observations: 139

AR Root(s)	Modulus	Cycle
0.194745	0.194745	

No root lies outside the unit circle.
 ARMA model is stationary.

MA Root(s)	Modulus	Cycle
0.979125	0.979125	

No root lies outside the unit circle.
 ARMA model is invertible.

obr.5.22: ARMA structure

Overíme heteroskedasticitu:

Heteroskedasticity Test: White

F-statistic	0.835728	Prob. F(14,123)	0.6294
Obs*R-squared	11.98682	Prob. Chi-Square(14)	0.6074
Scaled explained SS	15.56905	Prob. Chi-Square(14)	0.3404

obr.5.23: White test

Heteroskedasticita sa zamieta.

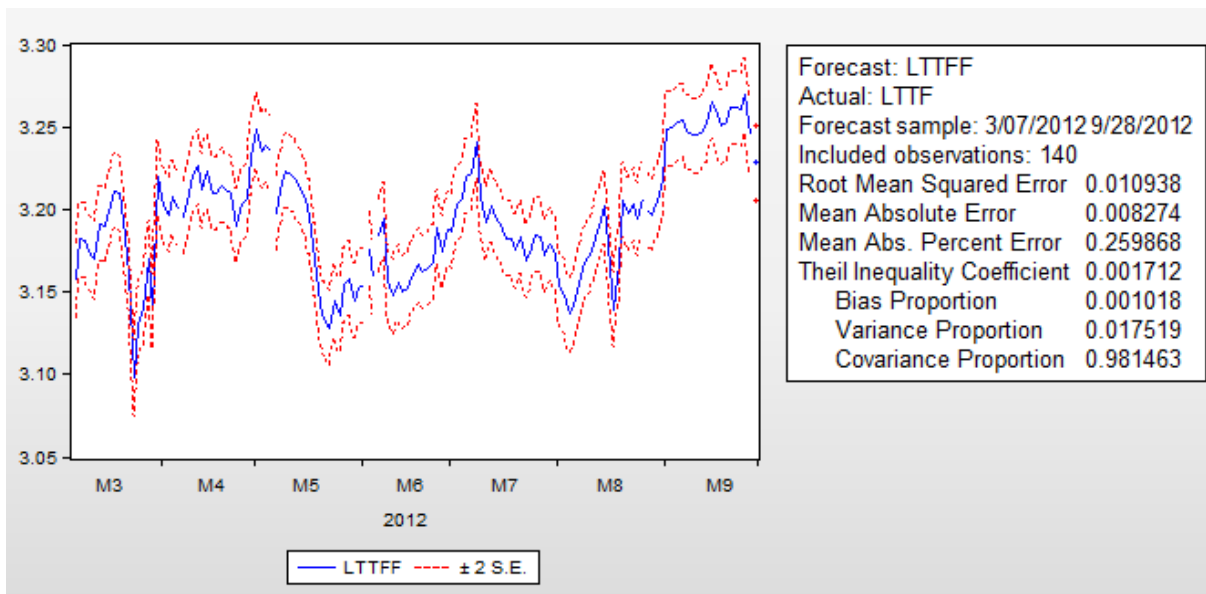
Durbin-Watsonova štatistika má hodnotu 2,002 takže nie je prítomná autokorelácia v reziduách. Potvrďme si to *Q-štatistikou*:

Date: 04/23/13 Time: 23:40
 Sample: 3/08/2012 9/28/2012
 Included observations: 139
 Q-statistic probabilities adjusted for 2 ARMA term(s)

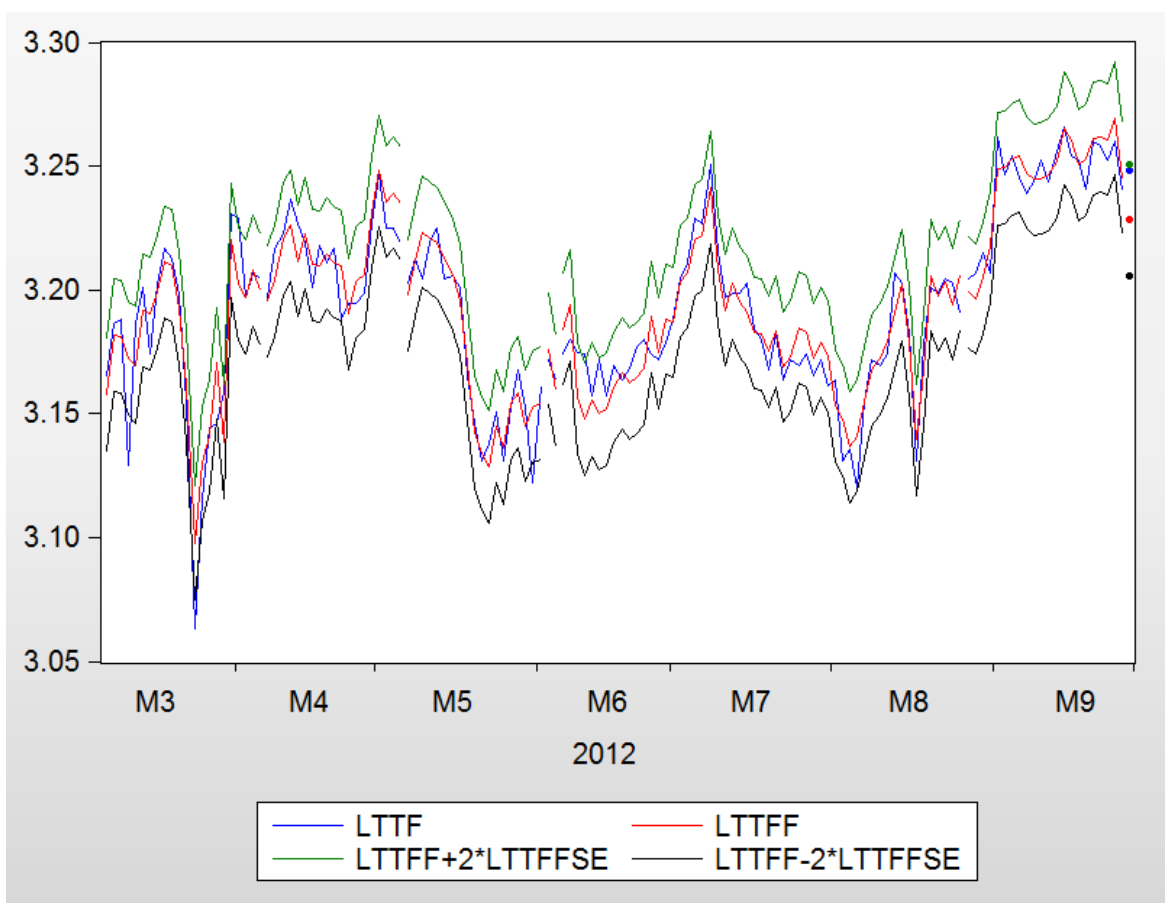
Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
		1	-0.017	-0.017	0.0389	
		2	0.058	0.058	0.5219	
		3	0.085	0.087	1.5685	0.210
		4	0.120	0.121	3.6657	0.160
		5	-0.052	-0.057	4.0609	0.255
		6	-0.026	-0.051	4.1608	0.385
		7	-0.067	-0.087	4.8263	0.437
		8	0.144	0.145	7.9217	0.244
		9	-0.024	0.014	8.0053	0.332
		10	-0.072	-0.073	8.7960	0.360
		11	-0.107	-0.134	10.538	0.309
		12	0.115	0.083	12.580	0.248
		13	-0.159	-0.118	16.500	0.124
		14	-0.089	-0.065	17.729	0.124
		15	-0.042	-0.015	18.013	0.157
		16	0.041	0.022	18.280	0.194
		17	-0.074	-0.036	19.171	0.206
		18	-0.048	-0.038	19.539	0.242
		19	-0.001	0.030	19.539	0.298
		20	0.030	-0.020	19.690	0.351
		21	-0.080	-0.055	20.764	0.350
		22	-0.032	-0.017	20.938	0.401
		23	-0.062	-0.056	21.587	0.424
		24	0.004	-0.056	21.590	0.485
		25	-0.095	-0.062	23.139	0.453
		26	-0.004	0.017	23.141	0.511
		27	-0.032	-0.039	23.324	0.559
		28	0.032	-0.007	23.500	0.605

obr.5.24: korelogram rezid.

Zo všetkých výstupov môžeme usúdiť, že skonštruovaný model je vyhovujúci. Prejdeme teda k predikciám a porovnáme ich s pozorovaniami.



obr.5.25:forecast TTF



obr.5.26 porovnanie predpovedí s pozorovaniami TTF

Výsledky predpovedí sú uspokojivé.

Záver

Zo skonštruovaných modelov môžeme vyvodiť záver že v prípade hubu BAGA sa zo všetkých ostatných trhov najviac prejavuje TTF a práve TTF je významne ovplyvnená trhom NBP. Toto zistenie korešponduje so skúsenosťami z praxe, ktoré nasvedčujú tomu, že vzájomný vplyv trhov smeruje zo západu na východ. Počas výskumu sme tvorili modely s rôznym časovým rozsahom a kombináciami exogénnych premenných. V tejto práci poukazujeme práve na modely BAGA a TTF, ktoré nám priniesli významné výsledky.

5.2 VAR modelovanie

V tejto časti vytvoríme VAR model, v ktorom budú všetky trhy zahrnuté ako endogénne premenné. Dĺžku oneskorení sme volili subjektívne s prihliadnutím na *Schwarzovo informačné kritérium*:

Vector Autoregression Estimates

Date: 04/24/13 Time: 00:24

Sample: 3/08/2012 9/28/2012

Included observations: 139

Standard errors in () & t-statistics in []

	LTF	LBAGA	LGASP	LNBP	LNCG
LTF(-1)	0.196506 (0.16373) [1.20021]	0.248392 (0.11969) [2.07530]	-0.091439 (0.12192) [-0.74999]	-0.067255 (0.17450) [-0.38541]	0.199713 (0.07230) [2.76211]
LTF(-2)	0.173214 (0.16815) [1.03011]	-0.371700 (0.12292) [-3.02380]	-0.079688 (0.12521) [-0.63641]	-0.136283 (0.17922) [-0.76043]	0.074289 (0.07426) [1.00041]
LBAGA(-1)	-0.278070 (0.11420) [-2.43493]	0.471422 (0.08348) [5.64680]	-0.273599 (0.08504) [-3.21729]	-0.263253 (0.12172) [-2.16282]	-0.018871 (0.05043) [-0.37418]
LBAGA(-2)	0.241163 (0.11553) [2.08737]	0.317818 (0.08446) [3.76293]	0.235017 (0.08603) [2.73169]	0.240782 (0.12314) [1.95537]	-0.004619 (0.05102) [-0.09053]
LGASP(-1)	0.486711 (0.19215) [2.53295]	0.073491 (0.14047) [0.52318]	0.858279 (0.14309) [5.99830]	0.475735 (0.20480) [2.32293]	0.493465 (0.08486) [5.81522]
LGASP(-2)	-0.263452 (0.21010) [-1.25392]	0.005362 (0.15359) [0.03491]	-0.161113 (0.15645) [-1.02978]	-0.046181 (0.22393) [-0.20623]	-0.233233 (0.09278) [-2.51369]
LNBP(-1)	0.397961 (0.14603) [2.72523]	-0.118488 (0.10675) [-1.10993]	0.221702 (0.10874) [2.03882]	0.724969 (0.15564) [4.65799]	0.104501 (0.06449) [1.62046]
LNBP(-2)	-0.163596 (0.14848) [-1.10178]	0.187685 (0.10855) [1.72907]	0.003236 (0.11057) [0.02927]	0.075075 (0.15826) [0.47439]	0.018190 (0.06557) [0.27741]
LNCG(-1)	0.125281 (0.19782) [0.63332]	-0.033233 (0.14461) [-0.22981]	0.109178 (0.14731) [0.74116]	0.027441 (0.21084) [0.13015]	0.255743 (0.08736) [2.92749]
LNCG(-2)	-0.097547 (0.10764) [-0.90625]	0.029811 (0.07869) [0.37886]	-0.045058 (0.08015) [-0.56215]	-0.088667 (0.11472) [-0.77289]	-0.004723 (0.04753) [-0.09936]
C	0.583445 (0.35514) [1.64287]	0.620315 (0.25962) [2.38933]	0.719852 (0.26446) [2.72202]	0.182202 (0.37851) [0.48136]	0.373315 (0.15683) [2.38031]

Pre zachovanie vzťahov medzi časovými radmi pracujeme s nediferencovanými dátami, pretože diferencovaním by sa mohli stratiť užitočné informácie.

Otestujeme stabilitu modelu:

Roots of Characteristic Polynomial
 Endogenous variables: LTTF LBAGA LGASP LNBP L...
 Exogenous variables: C
 Lag specification: 1 2
 Date: 04/24/13 Time: 00:34

Root	Modulus
0.883796	0.883796
0.838276	0.838276
0.592374	0.592374
0.475176	0.475176
-0.431410	0.431410
0.336421 - 0.089148i	0.348032
0.336421 + 0.089148i	0.348032
-0.223344 - 0.168236i	0.279618
-0.223344 + 0.168236i	0.279618
-0.077448	0.077448

No root lies outside the unit circle.
 VAR satisfies the stability condition.

VAR model je stabilný.

Pre hub BAGA dostávame rovnicu:

$$lbaga = 0.2483922 * lttf(-1) - 0.3716996 * lttf(-2) + 0.4714220 * lbaga(-1) + 0.3178175 * lbaga(-2) + 0.0734907 * lgasp(-1) + 0.0053618 * lgasp(-2) - 0.1184878 * lnbp(-1) + 0.1876850 * lnbp(-2) - 0.0332327 * lncg(-1) + 0.0298114 * lncg(-2) + 0.6203148$$

Pre hub TTF dostávame rovnicu:

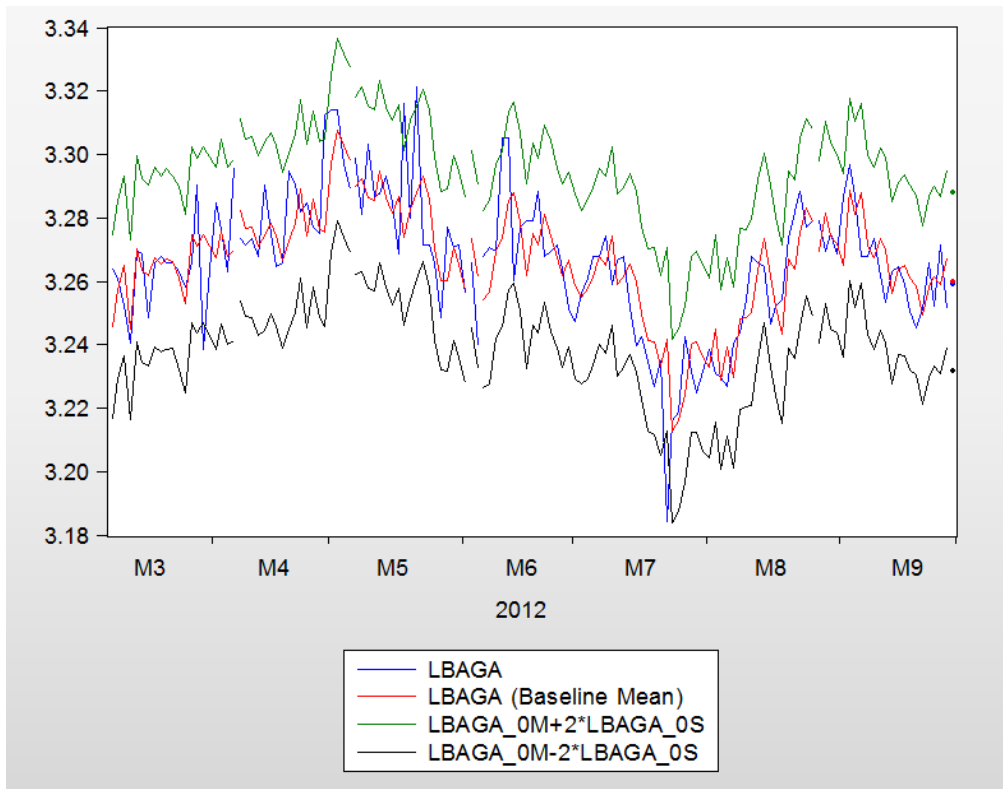
$$lttf = 0.1965064 * lttf(-1) + 0.1732136 * lttf(-2) - 0.2780701 * lbaga(-1) + 0.2411632 * lbaga(-2) + 0.4867110 * lgasp(-1) - 0.2634522 * lgasp(-2) + 0.3979607 * lnbp(-1) - 0.1635959 * lnbp(-2) + 0.1252808 * lncg(-1) - 0.0975469 * lncg(-2) + 0.5834446$$

Záver

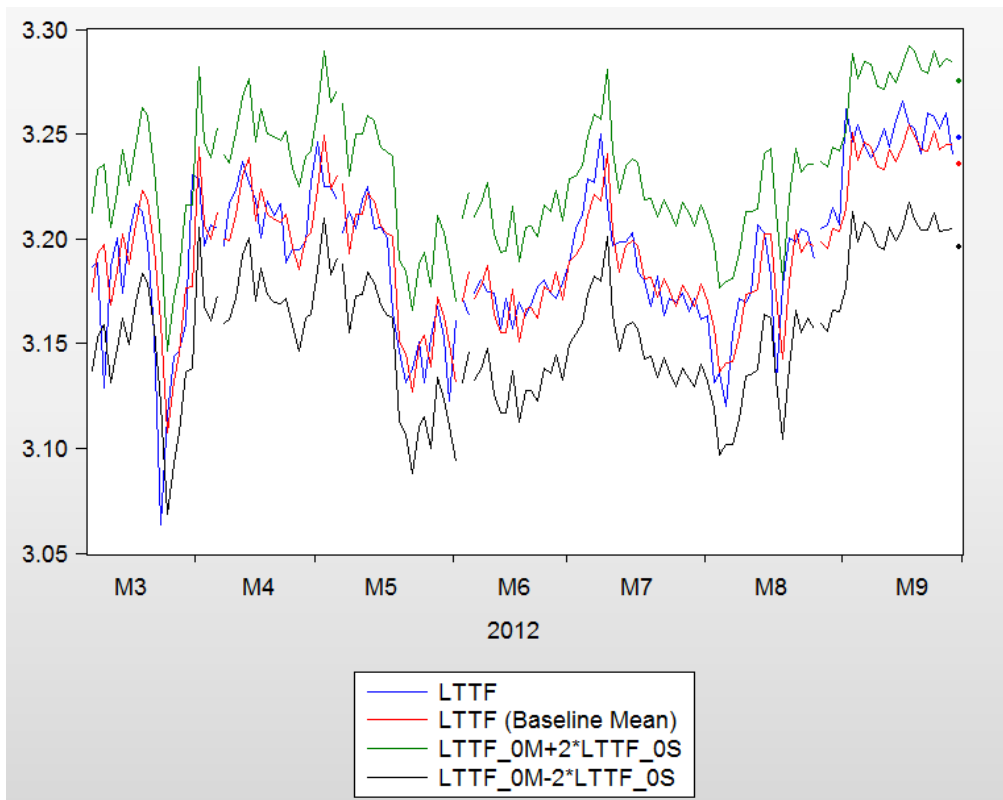
Vidno ze najvýznamnejšie koeficienty má BAGA pri vlastných hodnotách a pri hodnotách hubu TTF. Tento fakt sa zhoduje s výsledkom z ARIMA modelu, kde ako najvplyvnejšia exogénna premenná na BAGA pôsobila práve TTF.

V prípade TTF sa najvýznamnejšie koeficienty vyskytujú pri GASP a NBP. Môžeme teda súhlasiť s výsledkami ARIMA modelu pre TTF, že NBP má signifikantný vplyv na TTF.

Vykreslíme si predpovede VAR modelu pre BAGA a TTF. Pri riešení predpovedí sme zvolili statické riešenie so stochastickou simuláciou.

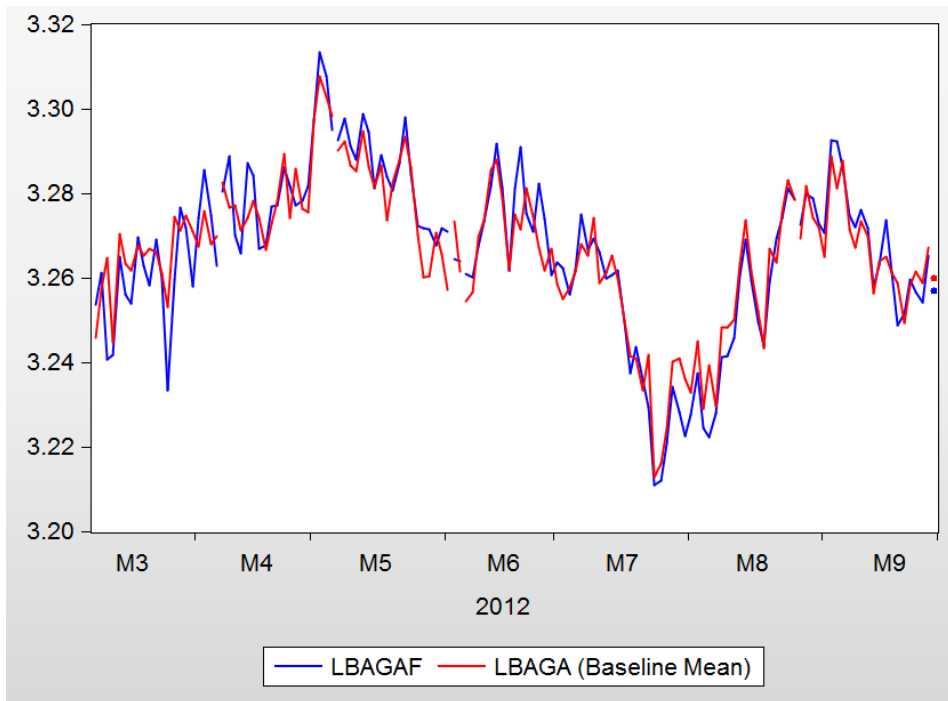


obr.5.27:VAR:BAGA

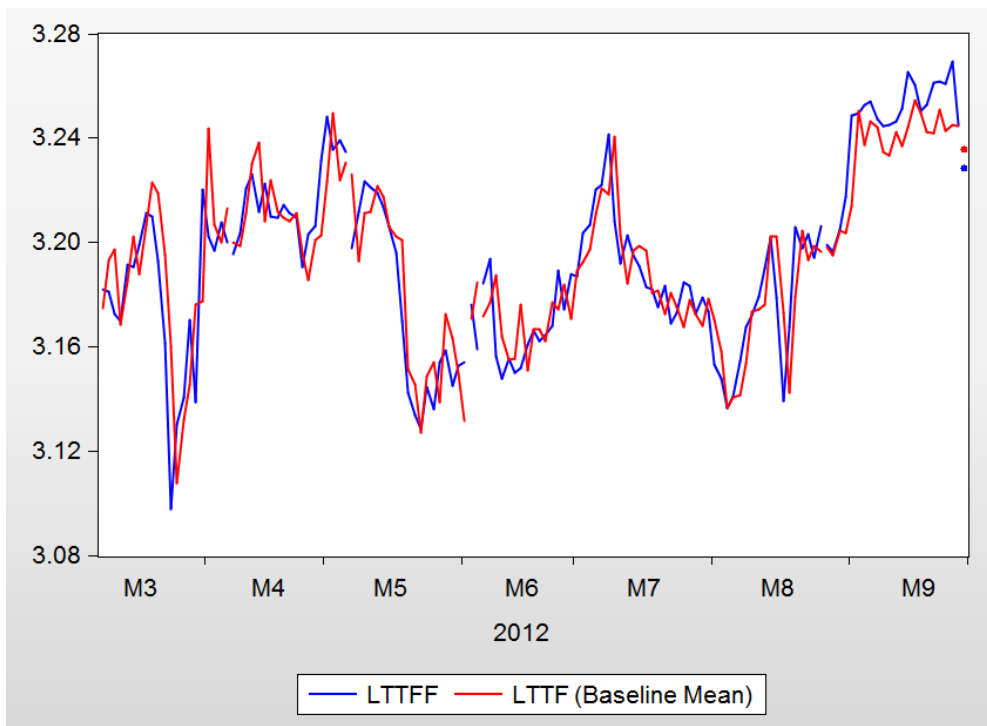


obr.5.28:VAR:TTF

Z grafov vidno, že predpovede získané z VAR modelu takmer kopírujú hodnoty pozorovaní s miernym omeškaním. Zobrazíme si ešte grafické porovnanie predikovaných hodnôt pomocou ARIMA a VAR modelov:



Obr.5.29: porovnanie predikcii ARIMA a VAR pre BAGA



Obr.5.30: porovnanie predikcii ARIMA a VAR pre TTF

5.3 Grangerova kauzalita

V tejto časti budeme skúmať vzájomný vplyv jednotlivých trhov. Aby sme predišli nožnej strate informácii o vzájomných vzťahoch, budeme pracovať s nediferencovanými dátami:

Pairwise Granger Causality Tests

Date: 04/24/13 Time: 01:29

Sample: 3/08/2012 9/28/2012

Lags: 2

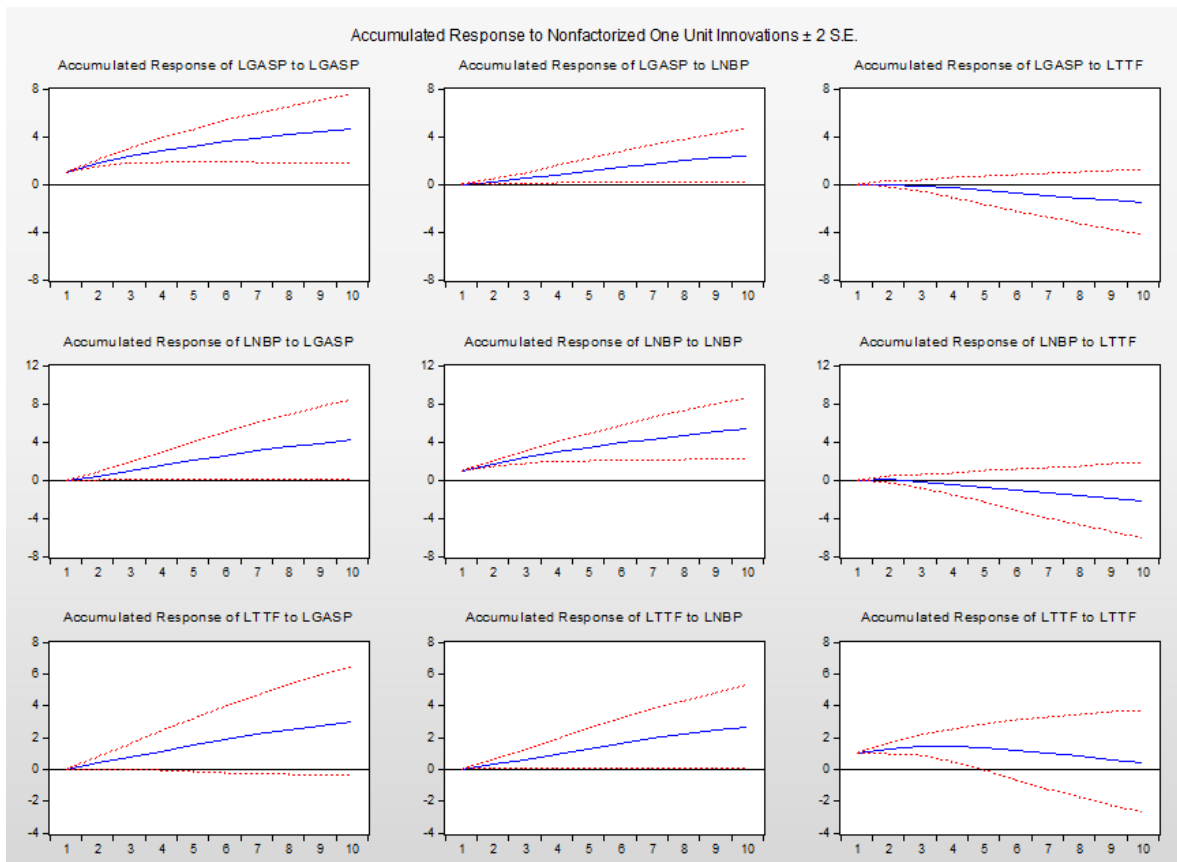
Null Hypothesis:	Obs	F-Statistic	Prob.
LBAGA does not Granger Cause LTTF	139	0.80819	0.4478
LTTF does not Granger Cause LBAGA		7.05629	0.0012
LGASP does not Granger Cause LTTF	139	4.66738	0.0110
LTTF does not Granger Cause LGASP		0.53351	0.5878
LNBP does not Granger Cause LTTF	139	5.44143	0.0053
LTTF does not Granger Cause LNBP		0.95424	0.3877
LNCG does not Granger Cause LTTF	139	0.16499	0.8481
LTTF does not Granger Cause LNCG		108.307	1.E-28
LGASP does not Granger Cause LBAGA	139	3.57576	0.0307
LBAGA does not Granger Cause LGASP		5.49513	0.0051
LNBP does not Granger Cause LBAGA	139	1.65285	0.1954
LBAGA does not Granger Cause LNBP		1.88655	0.1556
LNCG does not Granger Cause LBAGA	139	0.03210	0.9684
LBAGA does not Granger Cause LNCG		2.85717	0.0609
LNBP does not Granger Cause LGASP	139	2.18249	0.1168
LGASP does not Granger Cause LNBP		2.45849	0.0894
LNCG does not Granger Cause LGASP	139	0.17477	0.8398
LGASP does not Granger Cause LNCG		128.426	7.E-32
LNCG does not Granger Cause LNBP	139	0.14948	0.8613
LNBP does not Granger Cause LNCG		89.0774	2.E-25

Obr.5.29:Grangerova kauzalita

Z tohto výstupu sa dá vytvoriť pomyselný rebríček vzájomnej dominancie trhov na ktorého vrchole by bolo NBP nasledované GASP a TTF, ktoré značne ovplyvňujú NCG, ktoré je na podobnej úrovni ako BAGA.

5.4 Impulse response

Na záver si znázorníme reakcie trhov NBP,TTF a GASP na šoky, ktoré sa udejú na nejakom inom trhu :



Zobrazili sme si kumulatívne reakcie na jednotkovú zmenu rezídua. Šok na NBP spôsobí reakciu na GASP, ktorá sa na desiaty deň nakumuluje do zmeny o 2,4 jednotiek, v prípade TTF je to 2,7 jednotiek. Výrazné reakcie sú aj na šoky, ktoré vzniknú na GASP. TTF zareaguje kumulovanou zmenou o 3 jednotky a NBP dokonca až o 4,2 jednotky.

Záver

Cieľom tejto diplomovej práce bolo priblížiť čitateľovi základné charakteristiky časových radov a ich využitia pri tvorbe lineárnych modelov. Zamerali sme sa na jednorovnicové ARIMA modely a taktiež na vektorovo autoregresný VAR model. K dispozícii sme mali dáta denných cien zemného plynu na obchodných miestach (HUBoch) v Európe, cenu ropy a hodnotu výmenného kurzu USD/EUR pre obdobie 2010/1 až 2012/12. Pri tvorbe modelov sme sa snažili čo najviac využiť tieto dáta. Kvôli nepriaznivým výsledkom sme však obdobie pozorovaní zúžili na ceny počas obchodných dní v období 1.3.2012 až 28.9.2012, čím sme skúmali 144 pozorovaní. Taktiež sa nám nepodarilo dostatočne objasniť vplyv ropy na jednotlivé trhy, teda v práci sme sa venovali len vzájomnému vplyvu trhov.

Pomocou ARIMA modelov sme odhadli rovnice pre trhy BAGA a TTF, pri ktorých sa prejavil signifikantný vplyv zvyšných trhov. V prípade hubu BAGA sa zo všetkých ostatných trhov najviac prejavuje TTF a práve TTF je významne ovplyvnená trhom NBP. Toto zistenie korešponduje so skúsenosťami z praxe, ktoré nasvedčujú tomu, že vzájomný vplyv trhov smeruje zo západu na východ. Túto hypotézu sme overovali pomocou Grangerovej kauzality, ktorá ju sčasti potvrdzuje. Ak by sme si pospájali jednotlivé vplyvy, ktoré sú skúmané párovo, do jedného celku, ako najdominantnejší sa javí hub NBP. Za ním nasleduje GASP a TTF, ktoré významne ovplyvňujú NCG. Nemecký trh je teda ovplyvnený nemeckým a holandským trhom. Zaujímavým zistením je vzájomné pôsobenie GASP a BAGA.

Pomocou vektorovo autoregresného modelu sme taktiež vytvorili rovnice pre jednotlivé trhy, neskôr sme graficky porovnali predikcie ARIMA a VAR modelov.

Na záver sme stručne ukázali na vybraných troch huboch NBP, GASP a TTF ich reakcie na šoky vzniknuté na iných trhoch.

Návrh tradingovej stratégie spočíva vo využití rovníc pre jednotlivé trhy, ktoré sme získali zo skonštruovaných modelov, s prihliadnutím na signifikantné vzťahy medzi niektorými trhmi a taktiež na zohľadnenie reakcií na šoky.

Zoznam použitej literatúry

- [1] CIPRA, T., 1986, "Analyza časových řad s aplikacemi v ekonomii" SNTL - Nakladatelství technické literatury
- [2] LUKÁČIKOVÁ, A., LUKÁČIK, M., 2008, "Ekonometrické modelovanie s aplikáciami" Vydavateľstvo Ekonóm
- [3] SHUMWAY, R.H., and STOFFER, D.S., 2011, "Time Series Analysis and Its Applications" Springer
- [4] KOZÁK, J., HINDLS, R., ARLT, J., 1994, "Úvod do analýzy ekonomických časových řad" VŠE v Praze
- [5] KLÚČIK, M., 2010, "Základy programovacieho jazyka EViews a ich aplikácia na analýzy, prognózy a rýchle odhady vývoja makroekonomických ukazovateľov 1. Časť" Inštitút informatiky a štatistiky.
- [6] KLÚČIK, M., JURIOVÁ, J., 2011, "Základy programovacieho jazyka EViews a ich aplikácia na analýzy, prognózy a rýchle odhady vývoja makroekonomických ukazovateľov 2. Časť" Inštitút informatiky a štatistiky.
- [7] LUKÁČIK, M., 2012, "Štrukturálne vektorovo autoregresné modely"
- [8] FORBELSKÁ, M., "Stochastické modelování jednorozměrných časových řad" Masarykova univerzita, 2009
- [9] EViews 7 User's Guide I & II
- [10] ENDERS, W., 2004, "Applied Econometric Time Series, Second edition". John Wiley and sons.

Internetové zdroje:

<http://faculty.washington.edu/ezivot/introforecast.PDF>

<http://faculty.washington.edu/ezivot/econ584/notes/varModels.pdf>

<http://www.iam.fmph.uniba.sk/institute/stehlikova/cr09/cv1.html>

<http://www.iam.fmph.uniba.sk/institute/stehlikova/cr09/cv2.html>