Univerzita Karlova v Praze
Matematicko-fyzikální fakulta

BAKALÁŘSKÁ PRÁCE

Martin Čekal

Modelování sezónních ekonomických řad

Katedra pravděpodobnosti a matematické statistiky

Vedoucí bakalářské práce: RNDr. Jitka Zichová, Dr.
Studijní program: Matematika, Finanční matematika

2008

Velmi rád bych poděkoval vedoucí mé bakalářské práce RNDr. Jitce Zichové, Dr. za volbu zajímavého tématu, dále za cenné rady a připomínky při tvorbě práce a za poskytnutí studijní literatury.

Prohlašuji, že jsem svou bakalářskou práci napsal samostatně a výhradně s použitím citovaných pramenů. Souhlasím se zapůjčováním práce a jejím zveřejňováním.

V Praze dne 26.5.2008 Martin Čekal

Obsah

1 Úvod **6**

2 Principy modelování časových řad **7**
2.1 Dekompozice časových řad 7
2.2 Boxova - Jenkinsova metodologie 9
2.2.1 Aparát Boxovy – Jenkinsovy metodologie 9
2.2.2 Obecný lineární proces 12
2.2.3 Proces klouzavých součtů MA(q) 13
2.2.4 Autoregresní proces AR(p) 14
2.2.5 Smíšený proces ARMA(p,q) 15
2.2.6 Modely ARIMA(p,d,q) 16
2.2.7 Sezónní modely ARIMA 17
2.2.8 Konstrukce modelu 18

3 Výpočet v programu NCSS **20**
3.1 Dekompozice v programu NCSS 20
3.2 Exponenciální vyrovnávání se sezónní složkou 21
3.2.1 Aditivní sezónní složka 22
3.2.2 Multiplikativní sezónní složka 22
3.2.3 Konstanty vyrovnávání 22
3.3 Sezónní ARIMA v NCSS 23
3.3.1 Výstavba modelu 23
3.3.2 Specifikace modelu 24

4 Aplikace modelů na reálná data **26**
4.1 Data 1 26
4.1.1 Sezónní dekompozice 27
4.1.2 Exponenciální vyrovnávání 29
4.1.3 Sezónní ARIMA 34

4.2 Data 2 37
4.2.1 Sezónní dekompozice 38
4.2.2 Exponenciální vyrovnávání 41
4.2.3 Sezónní ARIMA 46
4.3 Shrnutí 48

Literatura **50**

Název práce: Modelování sezónních ekonomických řad
Autor: Martin Čekal
Katedra: Katedra pravděpodobnosti a matematické statistiky
Vedoucí bakalářské práce: RNDr. Jitka Zichová, Dr.
e-mail vedoucího: zichova@karlin.mff.cuni.cz

Abstrakt: V předložené práci studujeme některé přístupy k analýze časových řad a to především s důrazem na popis jejich sezónního chování. Rozebrány jsou postupně metody klasického deterministického přístupu jako *dekompozice* a *exponenciální vyrovnávání* a dále statistického přístupu *Boxovy - Jenkinsovy metodologie*, kam patří modely *AR*, *MA*, *ARMA*, *ARIMA* a *sezónní ARIMA*. Dále se práce zabývá jejich realizací ve statistickém programu NCSS a aplikací na konkrétní data z ekonomické praxe.

Klíčová slova: sezónní dekompozice, exponenciální vyrovnávání, smíšený proces ARMA, sezónní model ARIMA

Title: Modelling seasonal economical series
Author: Martin Čekal
Department: Department of Probability and Mathematical Statistics
Supervisor: RNDr. Jitka Zichová, Dr.
Supervisor's e-mail address: zichova@karlin.mff.cuni.cz

Abstract: In the present work we study some of the basic priciples in time series analysis, especially their seasonal behaviour. At first, this work studies deterministic models like *Decomposition* and *Exponential Smoothing*, further it concerns statistical models of *Box - Jenkins metodology* like *AR*, *MA*,*ARMA*, *ARIMA* and *Seasonal ARIMA*. Realization of these models is explained in statistical software NCSS. Finally, explained models are used in NCSS for modelling economical data.

Keywords: Seasonal Decomposition, Exponential Smoothing, ARMA model, ARIMA seasonal model

Kapitola 1

Úvod

Tato bakalářská práce rozebírá některé metody analýzy časových řad. Časovou řadou nazýváme posloupnost měření, uskutečněných v pravidelných časových intervalech, z různých oblastí. Nejčastěji se jedná o data z fyziky, techniky, biologie, společenských věd, ekologie nebo ekonomie. Analýza časových řad se zabývá konstrukcí vhodných modelů na popis těchto dat a jejich použití k předpovídání budoucího vývoje.

Časová řada může být *deterministická* nebo *statistická*. Pod pojmem *deterministická časová řada* rozumíme časovou řadu, jejíž chování lze striktně popsat matematickým vzorcem, bez použití náhodných veličin. U takto popsané časové řady lze zkonstruovat přesnou předpověď jejích budoucích hodnot.

Statistická časová řada je časová řada, jejíž chování obsahuje nejistotu, matematicky popsanou náhodnými veličinami. Předpovědi u takové časové řady jsou jen pravděpodobnostní.

Některé časové řady vykazují zjevný sezónní charakter, který je především z důvodu tvorby předpovědí potřeba v modelu zachytit. V následujících kapitolách budou postupně rozebrány některé modely spadající jak do kategorie deterministického, tak do kategorie statistického přístupu modelování časových řad a to s důrazem na popis jejich sezónního chování.

Kapitola 2

Principy modelování časových řad

2.1 Dekompozice časových řad

Některé časové řady mohou být z důvodu identifikace pravidelného chování rozloženy na několik složek:

Trend Tr_t,
Sezónní složka Sz_t,
Cyklická složka C_t,
Reziduální (náhodná) složka E_t.

Trend Tr_t odráží dlouhodobé změny v chování časové řady, např. technologické změny ve výrobě, změny v populaci, růst trhu, změny ve výši příjmů obyvatelstva. Má relativní charakter – např. změna klimatu se jeví zemědělci dlouhodobě, zato pro geologa se jedná o krátkodobý pohyb.

Sezónní složka Sz_t popisuje periodické změny v časové řadě odehrávající se během kalendářního roku a opakující se každý rok. Tyto změny jsou způsobeny především faktory jako střídání ročních období nebo sezónnost lidských zvyků. Příkladem mohou být změny v objemu měsíčního prodeje obchodního domu, změny v průměrných měsíčních teplotách, sezónnost počtu pasažérů leteckých společností atd. Pro měření této složky jsou vhodná měsíční měření, při méně častých měřeních se dá těžko popsat.

Cyklická složka C_t je nejspornější složkou časové řady. Někdy se dá považovat spíše za fluktuace kolem trendu, kde se střídá fáze růstu s fází poklesu. Může být výrazná i bez zjevných příčin. Pro západní ekonomiku je typickým představitelem této složky tzv. obchodní cyklus („business cycle“), který charakterizuje růst a pokles ekonomické aktivity. Jeho délka se pohybuje mezi 5 a 7 lety. Dále např. cyklické změny v klimatu, které dále způsobují cyklické změny v průtocích řek nebo zemědělské produkci. Cyklická složka se může objevit i tam, kde by to nikdo nečekal, jako např. v řadě tvořené cenovým indexem pšenice. Její průměrná délka je 13 let. Dále je významná v módě a oděvním průmyslu. Charakter této složky se může v čase měnit.

Reziduální složka E_t je náhodná část časové řady. Náhodné fluktuace nemají rozpoznatelný systematický charakter. Tato složka má několik funkcí: pokrývá chyby v měření časové řady a dále některé chyby, které vznikají při zpracovávání časové řady (např. zaokrouhlení údajů). Většinou se reziduální složka pokládá za bílý šum nebo bílý šum s normálním rozdělením, tedy posloupnost nekorelovaných náhodných veličin s nulovou střední hodnotou a konstantním rozptylem.

Časovou řadu lze rozložit do tvaru:

1) Aditivní dekompozice:

$$y_t = Tr_t + C_t + Sz_t + E_t \tag{2.1}$$

2) Multiplikativní dekompozice:

$$y_t = Tr_t * C_t * Sz_t * E_t \tag{2.2}$$

Při aditivním rozkladu jsou jednotlivé složky uvažovány ve svých absolutních hodnotách, tedy měřeny v jednotkách řady y(t), při multiplikativním rozkladu je většinou ve své absolutní hodnotě uvažována jen trendová složka a ostatní jsou v relativních hodnotách vůči trendu – tedy bezrozměrné . Po logaritmické transformaci přechází multiplikativní dekompozice na aditivní (navíc se však mění statistické vlastnosti bílého šumu).

Dekompoziční metody kladou důraz na práci se systematickými složkami časové řady. Tyto modely jsou deterministické. Do této kategorie patří

dekompozice pomocí klouzavých průměrů nebo různé typy *exponenciálního vyrovnávání*. Výpočet pomocí těchto modelů bude popsán v 2. kapitole.

2.2 Boxova - Jenkinsova metodologie

Jak jsme uvedli, časové řady založené na dekompozici jsou deterministické. Boxova - Jenkinsova metodologie zaujímá odlišný přístup k analýze časové řady. Jako základní prvek konstrukce modelu je brána reziduální složka a její vzájemné vztahy. Pohlížíme zde na celou řadu jako na náhodnou posloupnost s určitými speciálními vlastnostmi. Příkladem takové časové řady je:

$$y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1}, \tag{2.3}$$

kde θ_1 je parametr a ε_t, ε_{t-1} je bílý šum.

Mezi hlavní představitele této metodologie patří modely MA, autoregresní AR a smíšené ARMA. Dále ARIMA a sezónní ARIMA, které slouží k modelování řad s významnějším sezónním charakterem. Všechny tyto modely budou popsány níže. I přes výborné praktické výsledky mají zmíněné modely několik nevýhod:

1) Vyžadují relativně vysoký počet pozorování - je potřeba mít k dispozici řadu s alespoň 50 pozorováními, což je problém u řad s dlouhými intervaly mezi měřeními.

2) Interpretace výsledků není snadná.

2.2.1 Aparát Boxovy – Jenkinsovy metodologie

Stacionarita

Všeobecně řečeno stacionarita časové řady znamená, že chování řady je v jistém smyslu stochasticky ustálené. Rozlišuje se striktní stacionarita a slabá stacionarita (striktní stacionarita viz např.[3]). Slabá stacionarita znamená, že stochastický proces má konstantní střední hodnotu, konstantní rozptyl a

kovarianční strukturu druhého řádu invariantní vůči posunutí v čase, tedy:

$$cov(y_t, y_s) = cov(y_{t+h}, y_{s+h}) \tag{2.4}$$

pro libovolné h.

V Boxově – Jenkinsově metodologii lze modelovat pouze slabě stacionární (dále jen stacionární) řady, přičemž pomocí různých transformací lze mnoho nestacionárních řad převést na stacionární.

Autokovarianční a autokorelační funkce

Autokovarianční funkce γ_k je definována jako:

$$\gamma_k = cov(y_t, y_{t+k}) = E(y_t - \mu)(y_{t+k} - \mu), \quad k = ..., -1, 0, 1, ... \tag{2.5}$$

Autokorelační funkce ϱ_k je definována jako:

$$\varrho_k = \frac{\gamma_k}{\gamma_0} = \frac{\gamma_k}{\sigma_y^2}, \quad k = ..., -1, 0, 1, ... \tag{2.6}$$

Obě funkce jsou sudé, tedy pro jejich popis se stačí omezit na k $\geq$ 0. Vždy platí $|\varrho_k| \leq 1$ a $\varrho_0 = 1$. Grafický záznam autokorelační funkce se nazývá korelogram.

Odhad autokovarianční a autokorelační funkce

Označíme-li $y_1 ... y_n$ naměřené hodnoty časové řady, potom

odhadem střední hodnoty časové řady je

$$\overline{y} = \sum_{t=1}^{n} \frac{y_t}{n}, \tag{2.7}$$

odhadem autokovarianční funkce je

$$c_k = \sum_{t=1}^{n-k} \frac{(y_t - \overline{y})(y_{t+k} - \overline{y})}{n}, \quad k = 0, 1, ..., n-1. \tag{2.8}$$

Odhad autokorelační funkce:

$$r_k = \frac{c_k}{c_0}, k = 0, 1, ..., n-1. \tag{2.9}$$

Jak už bylo zmíněno výše, obvykle se požaduje $n \geq 50$, ale také k$\leq n/4$. Nutné je požadovat také ergodicitu příslušného procesu (viz [2]). Jedná se o asymptoticky nestranné odhady, jelikož

$$E(c_k) \overset{n\to\infty}{\to} \gamma_k. \tag{2.10}$$

Pro manipulaci s modelem je důležité umět spočítat hodnotu k_0 takovou, že hodnota $\rho_k = 0$ pro $k > k_0$, nebo zjistit, že taková hodnota neexistuje. Je tedy třeba vyšetřit, jak blízko nule musí být r_k, abychom mohli ϱ_k prohlásit za rovné nule. Spočítáme směrodatnou odchylku odhadu r_k za podmínky, že $\varrho_k = 0$ pro $k > k_0$. K tomu použijeme za předpokladu normality procesu Bartlettovu aproximaci

$$\sigma(r_k) = \sqrt{var(r_k)} \sim \sqrt{\frac{1}{n}(1 + 2\sum_{j=1}^{k_0} r_j^2)}, \quad k > k_0. \tag{2.11}$$

Vzhledem k tomu, že $P(|X| > 2\sigma) \doteq 0.05$, kde $X \sim N(0, \sigma^2)$, můžeme pak rozhodnout, že $\varrho_k = 0$ když $|r_k| < 2\sigma(r_k)$.

Parciální autokorelační funkce

Parciální autokorelační funkce ϱ_{kk} je definována jako parciální korelační koeficient y_t a y_{t+k} při pevných hodnotách $y_{t+1}, ..., y_{t+k-1}$. Platí pro ni vztah

$$\varrho_{kk} = \frac{|P_k^*|}{|P_k|}, \tag{2.12}$$

kde $|\ |$ je determinant matice a

$$|P_k| = \begin{vmatrix} 1 & \varrho_1 & \cdots & \varrho_{k-1} \\ \varrho_1 & 1 & \cdots & \varrho_{k-2} \\ \vdots & \vdots & \ddots & \vdots \\ \varrho_{k-1} & \varrho_{k-2} & \cdots & 1 \end{vmatrix} \tag{2.13}$$

a

$$|P_k^*| = \begin{vmatrix} 1 & \varrho_1 & \cdots & \varrho_1 \\ \varrho_1 & 1 & \cdots & \varrho_2 \\ \vdots & \vdots & \ddots & \vdots \\ \varrho_{k-1} & \varrho_{k-2} & \cdots & \varrho_k \end{vmatrix} \tag{2.14}$$

Odhady r_{kk} parciální autokorelační funkce se v počítačových programech obvykle počítají rekurentně. U parciální autokorelační funkce se obvykle testuje nulovost hodnot na základě Quenouilleovy aproximace: Je-li $\varrho_{kk} = 0$ pro $k > k_0$, pak

$$\sigma(r_{kk}) \sim \sqrt{\frac{1}{n}}, \quad k > k_0. \tag{2.15}$$

Pokud je dvojnásobek této hodnoty číslem $|r_{kk}|$ překročen, lze hodnotu parciální autokorelační funkce považovat za nenulovou.

2.2.2 Obecný lineární proces

Lineární proces je řada tvaru:

$$y_t = \varepsilon_t + \psi_1 \varepsilon_{t-1} + \psi_2 \varepsilon_{t-2} + \ldots, \tag{2.16}$$

kde ε_t je bílý šum a ψ_j jsou parametry. Jestliže definujeme tzv. operátor zpětného posunutí B jako

$$By_t = y_{t-1} \tag{2.17}$$

$$B^j y_t = y_{t-j}, \tag{2.18}$$

platí

$$y_t = \psi(B)\varepsilon_t, \tag{2.19}$$

kde

$$\psi(B) = 1 + \psi_1 B + \psi_2 B^2 + \ldots = 1 + \sum_{j=1}^{\infty} \psi_j B^j. \tag{2.20}$$

Jedná se tedy o mocninou řadu v proměnné B. Pro existenci lineárního procesu musí být splněno, že $\psi(B)$ konverguje pro $|B| \leq 1$. Tato podmínka dále zaručí stacionaritu lineárního procesu a také $E(y_t) = 0$.

Jestliže lineární proces splňuje navíc podmínku, že

$$\pi(B) = 1 - \pi_1 B - \pi_2 B^2 - \ldots = 1 - \sum_{j=1}^{\infty} \pi_j B^j, \tag{2.21}$$

konverguje pro $|B| \leq 1$, lze jej zapsat do následujícího tvaru:

$$y_t = \pi_1 y_{t-1} + \pi_2 y_{t-2} + \ldots + \varepsilon_t, \tag{2.22}$$

nebo také

$$\pi(B) y_t = \varepsilon_t, \tag{2.23}$$

přičemž B je opět operátor zpětného posunutí.
Pokud lze proces zapsat v tomto tvaru, pak se nazývá *invertibilní*. Pro parametry ψ_j a π_j platí rovnost:

$$\psi(B)\pi(B) = 1. \tag{2.24}$$

Dále se budeme zabývat speciálními případy lineárního procesu.

2.2.3 Proces klouzavých součtů MA(q)

Proces klouzavých součtů řádu q má tvar:

$$y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + ... + \theta_q \varepsilon_{t-q}, \tag{2.25}$$

neboli

$$y_t = \theta(B)\varepsilon_t, \tag{2.26}$$

kde

$$\theta(B) = 1 + \sum_{j=1}^{q} \theta_j B^j, \tag{2.27}$$

což je tzv. operátor klouzavých součtů.

Proces MA(q) je pro každou volbu parametrů stacionární podle podmínky pro obecný lineární proces, střední hodnota je nulová a rozptyl je:

$$\sigma_y^2 = \gamma_0 = (1 + \theta_1^2 + ... + \theta_q^2)\sigma_\epsilon^2 \tag{2.28}$$

Proces je invertibilní, jestliže řada $\pi(B) = \theta^{-1}(B)$ konverguje pro $|B| \leq 1$. Jsou-li $H_1, ..., H_q$ kořeny polynomu $\theta(B)$, pak lze tento polynom rozložit na

$$\theta(B) = \prod_{j=1}^{q} (1 - \frac{B}{H_j})$$

a konvergenci řady $\theta^{-1}(B)$ pro $|B| \leq 1$ zaručí

$$|H_j| > 1 \quad j = 1, ..., q.$$

Tedy proces MA(q) je invertibilní, jestliže všechny kořeny polynomu $\theta(B)$ leží vně jednotkového kruhu v komplexní rovině.

Autokorelační funkce procesu MA(q):

$$\rho_k = \frac{\theta_k + \theta_1\theta_{k+1} + ... + \theta_{q-k}\theta q}{1 + \theta_1^2 + ... + \theta_q^2} \tag{2.29}$$

pro $k \leq q$ a

$$\rho_k = 0 \tag{2.30}$$

pro $k > q$, zřejmě tedy $k_0 = q$.

Řešením této nelineární soustavy rovnic při daných hodnotách ρ_k lze iteračním procesem získat hodnoty parametrů $\theta_1, ..., \theta_q$.

Parciální autokorelační funkce je omezená *křivkou U*, což je lineární kombinace geometricky klesajících posloupností a sinusoid různých frekvencí s geometricky klesajícími amplitudami. Identifikační bod k_0 pro ni neexistuje.

2.2.4 Autoregresní proces AR(p)

Jedná se o proces řádu p a má tvar:

$$y_t = \varphi_1 y_{t-1} + ... + \varphi_p y_{t-p} + \varepsilon_t, \tag{2.31}$$

neboli

$$\varphi(B)y_t = \varepsilon_t, \tag{2.32}$$

kde

$$\varphi(B) = 1 - \sum_{j=1}^{p} \varphi_j B^j. \tag{2.33}$$

Jedná se tedy o obecný lineární proces zapsaný v invertované formě, tedy je bez jakýchkoli podmínek invertibilní. Proces AR(p) je stacionární, jestliže všechny kořeny polynomu $\varphi(B)$ leží vně jednotkového kruhu v komplexní rovině. Střední hodnota je nulová a autokorelační funkce splňuje soustavu:

$$\varrho_k = \varphi_1 \varrho_{k-1} + \varphi_2 \varrho_{k-2} + ... + \varphi_p \varrho_{k-p}, \quad k = 1, ..., p. \tag{2.34}$$

Toto je Yuleova-Walkerova soustava rovnic, jejímž řešením dostáváme parametry $\varphi_1, ..., \varphi_p$ vyjádřené pomocí $\varrho_1, ..., \varrho_p$. Dá se ukázat, že rozptyl procesu AR(p) je:

$$\sigma_y^2 = \gamma_0 = \frac{\sigma_\varepsilon^2}{1 - \varphi_1\varrho_1 - ... - \varphi_p\varrho_p} \tag{2.35}$$

Pro parciální autokorelační funkci procesu AR(p) je $k_0 = p$. Autokorelační funkce má tvar křivky U.

2.2.5 Smíšený proces ARMA(p,q)

Jedná se o sloučení procesů AR(p) a MA(q). Je tedy tvaru

$$y_t = \varphi_1 y_{t-1} + ... + \varphi_p y_{t-p} + \varepsilon_t + \theta_1\varepsilon_{t-1} + ... + \theta_q\varepsilon_{t-q}, \tag{2.36}$$

neboli

$$\varphi(B)y_t = \theta(B)\varepsilon_t. \tag{2.37}$$

Podmínka stacionarity je totožná s podmínkou stacionarity AR(p). Podmínka invertibility je totožná s podmínkou invertibility MA(q). Střední hodnota stacionárního procesu je nulová. Autokorelační funkce procesu ARMA splňuje stejnou soustavu jako autokorelační funkce procesu AR(p) s rozdílem, že index k probíhá jen od q do p. Tedy:

$$\varrho_k = \varphi_1\varrho_{k-1} + \varphi_2\varrho_{k-2} + ... + \varphi_p\varrho_{k-p}, \quad k > q. \tag{2.38}$$

Parciální autokorelační funkce ϱ_{kk} se chová stejně jako v procesu MA(q), ale pro $k \geq max(1, p - q + 1)$, neexistuje tedy pro ni identifikační bod k_0. Shrnutí je v následující tabulce:

Tabulka 2.1: Vlastnosti autokorelačních a parciálních autokorelačních funkcí u procesů AR, MA a ARMA.

	AR(p)	**MA(q)**	**ARMA(p,q)**
ϱ_k	neexistuje k_0 ϱ_k ve tvaru U	$k_0 = q$	neexistuje k_0, ϱ_k ve tvaru U po prvních $q - p$ hodnotách
ϱ_{kk}	$k_0 = p$	neexistuje k_0 ϱ_{kk} omezená křivkou U	neexistuje k_0, ϱ_{kk} omezená křivkou U po prvních $p - q$ hodnotách

Křivka U značí lineární kombinaci geometricky klesajících posloupností a sinusoid s geometricky klesající amplitudou.

2.2.6 Modely ARIMA(p,d,q)

Modely uvedené dosud předpokládaly stacionaritu časové řady. Pomocí modelů ARIMA můžeme však modelovat i řady, které mají náhodné změny ve velikosti nebo sklonu úrovně. Modelují totiž náhodně i trendovou složku. Tohoto je docíleno pomocí práce s diferencovanou řadou a to většinou pomocí první nebo druhé diference. Na diferencovanou řadu se poté uplatní výše zmíněné modely ARMA.

Při konstrukci modelu ARIMA tedy nepožadujeme stacionaritu řady, ale řada musí být převoditelná na stacionární pomocí diferencí. Pokud řada vykazuje stacionaritu až na náhodné změny ve velikosti úrovně, uplatní se diference 1. řádu. Pokud vykazuje stacionaritu až na náhodné změny ve sklonu úrovně, uplatní se diference 2. řádu.

Tedy model ARIMA(p,d,q) má tvar:

$$\varphi(B)w_t = \theta(B)\varepsilon_t, \tag{2.39}$$

kde

$$w_t = \Delta^d y_t \tag{2.40}$$

je d-tá diference modelovaného procesu y_t a (2.39) je stacionární model ARMA pro proces w_t.

Diferenční operátor Δ lze vyjádřit pomocí operátoru zpětného posunutí B jako:

$$\Delta = 1 - B, \tag{2.41}$$

protože $\Delta y_t = y_t - y_{t-1} = (1 - B)y_t$. Platí tedy také $\Delta^2 y_t = (1 - B)^2 y_t = (1 - 2B + B^2)y_t = y_t - 2y_{t-1} + y_{t-2}$.

Model ARIMA(p,d,q) můžeme tedy psát také jako:

$$\varphi(B)(1 - B)^d y_t = \theta(B)\varepsilon_t. \tag{2.42}$$

Operátor

$$v(B) = \varphi(B)(1 - B)^d$$

se nazývá zobecněný autoregresní operátor. Jedná se o polynom řádu p + d v proměnné B, který má s ohledem na stacionaritu (2.39) právě p kořenů vně jednotkového kruhu v komplexní rovině a d kořenů rovných jedné.

Diferencováním do řádu d se řada zkrátí z $y_1, ..., y_n$ na $w_{d+1}, .., w_n$. Také je důležité si uvědomit, že $\Delta^d(y_t - \overline{y}) = \Delta^d y_t$, tedy nemá smysl řadu centrovat.

Dá se ukázat, že model ARIMA dokáže kromě stochastického modelování úspěšně modelovat i čistě deterministické trendy.

2.2.7 Sezónní modely ARIMA

Stejně jako trendová složka se v modelech Boxovy-Jenkinsovy metodologie modeluje sezónní složka stochasticky. Mějme řadu měsíčních měření (tedy s periodou L = 12). Nejprve zkonstruujeme řadu např. lednových měření:

$$\Phi(B^{12})\Delta_{12}^D y_t = \Theta(B^{12})\eta_t, \tag{2.43}$$

t odpovídá lednovým obdobím a

$$\Phi(B^{12}) = 1 - \Phi_1 B^{12} - \Phi_2 B^{24} - ... - \Phi_P B^{P*12} \tag{2.44}$$

je sezónní autoregresní operátor řádu P,

$$\Theta(B^{12}) = 1 + \Theta_1 B^{12} + \Theta_2 B^{24} + ... + \Theta_Q B^{12*Q} \tag{2.45}$$

je sezónní operátor klouzavých součtů řádu Q a

$$\Delta_{12} = 1 - B^{12} \tag{2.46}$$

je sezónní diferenční operátor.

Podobné modely sestrojíme pro ostatní měsíce. Náhodné složky η_t budou mezi sebou korelované, jelikož existuje vztah mezi např. lednovými a únorovými měřeními. Nyní popíšeme tuto řadu η_t také modelem ARIMA:

$$\varphi(B)\Delta^d \eta_t = \theta(B)\varepsilon_t. \tag{2.47}$$

Zde ε_t již opravdu představuje bílý šum. Spojíme-li modely (2.47) a (2.43) dostaneme:

$$\varphi(B)\phi(B^{12})\Delta^d \Delta_{12}^D y_t = \theta(B)\Theta(B^{12})\varepsilon_t. \tag{2.48}$$

Tento model se značí SARIMA.

2.2.8 Konstrukce modelu

Konstrukce modelu pro stacionární řadu probíhá ve třech fázích. Identifikace modelu, odhad parametrů modelu a ověřování modelu.

Identifikace modelu - v této fázi se rozhodujeme o výběru vhodného modelu. Na výběr máme AR, MA, ARMA, případně ARIMA. Nejprve řadu centrujeme, tedy provedeme test o nulovosti její střední hodnoty, pokud je nenulová, odečteme odhad střední hodnoty.

Identifikace probíhá na základě vyšetření odhadů r_k a r_{kk}. U těchto funkcí se snažíme nalézt bod k_0. Podle tabulky 2.1 v (2.2.5) vybereme příslušný model.

Parametr q u modelu MA(q) určíme jako pořadí poslední významné hodnoty r_k. Pokud taková hodnota k_0 neexistuje, jedná se pravděpodobně o model AR(p), tedy q=0, nebo o model ARIMA(p,d,q) s kladnými p,d.

Parametr p u modelu AR(p) určíme jako pořadí poslední významné hodnoty r_{kk}. Pokud taková hodnota k_0 neexistuje, jedná se pravděpodobně o model MA(q), tedy p=0, nebo o model ARIMA(p,d,q) s kladnými q,d.

Jestliže hodnota k_0 neexistuje pro r_k ani r_{kk}, jedná se o model ARIMA a jeho autokorelační funkce bude tvaru křivky U, tedy lineární kombinace geometricky klesajících posloupností a sinusoid s geometricky klesající amplitudou po q-p hodnotách. Parciální autokorelační funkce bude mít podobný tvar po p-q hodnotách.

U modelu ARIMA se hodnoty p a q určují relativně obtížně, zde přichází na řadu zkoušení různých kombinací parametrů a posouzení residuí.

Odhad parametrů - počítačové programy mají vložené vzorce pro odhady hodnot parametrů $\varphi, \Phi, \theta, \Theta$ z hodnot r_k a r_{kk} a to podle toho, který model jsme vybrali v identifikační části. Poté se odhady zpřesňují vnitřní iterační odhadovou procedurou programu.

Ověřování modelu - model se ověřuje testováním nekorelovanosti residuí. Tento předpoklad je nezbytný pro použití iteračních procesů při stanovování hodnot parametrů θ a φ. K testování nekorelovanosti residuí slouží

např. různé statistické testy nebo posouzení histogramu. Pokud jsou korelace významné, považuje se model za neadekvátní. Dále o hodnověrnosti modelu může vypovídat hodnota MSE, což je průměr kvadrátů residuí. $MSE = \frac{1}{n}\sum_{t=1}^{n} e_t^2$.

Kapitola 3

Výpočet v programu NCSS

3.1 Dekompozice v programu NCSS

Program NCSS má vestavěnou jen multiplikativní dekompozici časové řady. Model tedy vypadá:

$$X_t = U \cdot Tr_t \cdot S_t \cdot C_t \cdot R_t,$$

kde U je průměr, Tr_t trend, S_t sezónní složka, C_t cyklická složka, R_t náhodná složka. Dekompozice se konstruuje následujícím způsobem:

1) Vydělení průměrem: Každá hodnota řady je vydělena průměrem. Tím vzniká nová řada $Y_t = \frac{X_t}{U}$. Je-li průměr menší než 10^{-7}, tento krok se vynechává.

2) Výpočet klouzavých průměrů: Mějme délku sezóny L. Nejčastěji $L = 12$. Spočítáme-li klouzavé průměry s poloměrem poloviny roku, odstraníme tím sezónní složku z řady. L může být sudé nebo liché. V případě, že je liché, není výpočet ničím komplikován a hodnoty se průměrují od $t - [L/2]$ do $t + [L/2]$, kde [x] značí celou část čísla x. Tedy

$$M_t = \frac{1}{L} \sum_{t-[L/2]}^{t+[L/2]} Y_t.$$

Je-li L sudé, je výpočet komplikovanější. Zprůměrovali bychom 12 hodnot, střed by byl v bodě 6,5. Tedy chceme-li spočítat klouzavý průměr v bodě 7, spočteme průměr v bodech 6,5 a 7,5 a tyto spolu zprůměrujeme.

Na koncích řady nastává problém při výpočtu průměrů vzhledem k neznalosti dat za nimi. Mějme např. řadu měření od ledna 2000 do prosince 2004. K výpočtu klouzavých průměrů bychom potřebovali znát např. data z prosince 1999. Ta se odhadnou tak, že se rozdíl v datech z prosince 2000 a prosince 2001 odečte od prosince 2000. Tedy prosinec 1999 = prosinec 2000 − (prosinec 2001 − prosinec 2000). Stejně se postupuje při "odhadování" dat za koncem časové řady, kde hodnota leden 2005 = leden 2004 + (leden 2004 - leden 2003). Takto vzniklou řadu označíme M_t.

3) Výpočet trendu: Použitím metody nejmenších čtverců (minimalizujeme ε_t) na model $M_t = a + b \cdot t + \varepsilon_t$ zjistíme parametry a a b, kde a je posunutí, b je směrnice přímky trendu a ε_t je residuum. Vzniká trend $Tr_t = a + bt$.

4) Výpočet cyklické složky: $C_t = \frac{M_t}{Tr_t}$.

5) Výpočet sezónní složky: Sezónní řada s residuem vznikne vydělením řady Y_t řadou klouzavých průměrů M_t. $K_t = \frac{Y_t}{M_t}$. Ještě zbývá oddělit residuum, to se provede tak, že se spočítá průměr všech pozorování ve stejné fázi sezóny, tedy např. průměr všech lednových měření. Tím vzniká sezónní řada S_t očištěná od residua. $S_t = \sum_{i=1}^{n} K_t$, n je počet sezón v řadě.

6) Výpočet residuí: $R_t = \frac{K_t}{S_t}$.

7) Tvorba předpovědí: Konstrukce předpovědi je velice jednoduchá. Trendová složka je $Tr_t = a + bt$, sezónní složka je $S_t = \sum_{i=1}^{n} K_t$, cyklická složka je zadána uživatelem nebo položena rovna 1 a residuum je 1. Pokud byla použita logaritmická transformace, jsou předpovědi transformovány zpět inverzní operací.

3.2 Exponenciální vyrovnávání se sezónní složkou

Tato metoda modeluje sezónní časové řady s trendem. Program NCSS zde používá Holt-Wintersův algoritmus exponenciálního vyrovnávání. Sezónní složka může být modelována aditivně nebo multiplikativně. Mějme $X_1, ..., X_n$ naměřená data časové řady.

3.2.1 Aditivní sezónní složka

Algoritmus modeluje časovou řadu tak, že se sezónní složka vyvíjí kolem trendu aditivně a používá pro předpovědi tyto vzorce:

$$a_t = \alpha(X_t - F_{t-s}) + (1-\alpha)(a_{t-1} + b_{t-1}) \tag{3.1}$$

$$b_t = \beta(a_t - a_{t-1}) + (1-\beta)b_{t-1} \tag{3.2}$$

$$F_t = \gamma(X_t - a_t) + (1-\gamma)F_{t-s}, \tag{3.3}$$

kde α, β, γ jsou konstanty vyrovnání, jejich hodnoty jsou mezi 0 a 1. a_t je posunutí a b_t je směrnice v čase t, s je počet period během roku nebo jiné sezóny. Předpověď hodnoty v čase $t+k$ z hodnoty v čase t spočteme jako $a_t + b_t k + F_{[(t+k-1)/s]+1}$. Zde $[(t+k-1)/s]$ značí zbytek po dělení $t+k-1$ číslem s, tedy dostaneme sezónu, ve které se $t+k$ nachází.

3.2.2 Multiplikativní sezónní složka

Zde NCSS používá Holt-Wintersův multiplikativní algoritmus, tedy modelujeme řadu, jejíž sezónní složka se vyvíjí multiplikativně kolem trendu. Tento model se hodí na ekonomické časové řady více než aditivní. Předpovědi se počítají pomocí těchto vzorců:

$$a_t = \alpha(X_t/F_{t-s}) + (1-\alpha)(a_{t-1} + b_{t-1}) \tag{3.4}$$

$$b_t = \beta(a_t - a_{t-1}) + (1-\beta)b_{t-1} \tag{3.5}$$

$$F_t = \gamma(X_t/a_t) + (1-\gamma)F_{t-s}, \tag{3.6}$$

kde α, β, γ jsou konstanty vyrovnání, jejich hodnoty jsou mezi 0 a 1. a_t je posunutí a b_t je směrnice v čase t, s je počet period během roku nebo jiné sezóny. Předpověď hodnoty v čase $t+k$ z hodnoty v čase t spočteme jako $(a_t + b_t k)F_{[(t+k-1)/s]+1}$.

3.2.3 Konstanty vyrovnávání

Tyto konstanty ovlivňují váhu minulých pozorování, tedy čím blíže jsou 1, tím více se při výpočtu předpovědí klade důraz na nedávné hodnoty, zato při zmenšování těchto konstant se zvětšuje vliv vzdálenějších hodnot. α representuje úroveň řady, β trend a γ sezónní složku. Volba těchto konstant závisí na uživateli programu, dá se odhadnout ze zkušenosti, nebo NCSS

nabízí 3 metody jejich iterace: MSE (mean square error), MAE (mean absolute error), MAPE (mean absolute percent error). Chyby, které jsou volbou parametrů minimalizovány, jsou:

$$e_t = X_t - F_{t-1}. \tag{3.7}$$

Jedná se tedy o rozdíl předpovídané hodnoty v čase t z času $t-1$ a skutečné hodnoty. Platí:

$$MSE = \frac{1}{n}\sum_{t=1}^{n} e_t^2 \tag{3.8}$$

$$MAE = \frac{1}{n}\sum_{t=1}^{n} |e_t| \tag{3.9}$$

$$MAPE = \frac{100}{n}\sum_{t=1}^{n} |\frac{e_t}{X_t}|. \tag{3.10}$$

Pro nalezení parametrů je třeba vybrat jedno z kritérií a NCSS nalezne parametry α a β pomocí *efficient grid-searching* algoritmu. Předtím než dojde ke spuštění Holt-Wintersova algoritmu, musíme znát hodnotu předpovědi v čase 0. Dále bychom potřebovali znát hodnoty prvních s sezónních faktorů. K tomuto účelu se používá metoda backcasting. Ta spočívá v obrácení časové řady a předpovídání těchto počátečních hodnot. Poté se řada opět obrátí zpět a předpověď může začít.

3.3 Sezónní ARIMA v NCSS

3.3.1 Výstavba modelu

Identifikace

Identifikace modelu probíhá v případě volby *Automatic ARMA* automaticky, v případě volby *Box-Jenkins ARIMA* musíme sami identifikovat nejvhodnější volbu počtu parametrů a typ modelu, jak je popsáno v předchozí kapitole. Volba počtu parametrů p, P, q, Q, d a D závisí tedy na tvaru autokorelační a parciální autokorelační funkce. Často v tomto případě vybíráme takový model, abychom minimalizovali MSE a zachovali nekorelovanost residuí.

Odhad parametrů

Program NCSS využívá k tomuto účelu vnitřní iterační procedury, které jsou modifikací *podmíněné* nebo *nepodmíněné odhadové metody nejmenších nelineárních čtverců.*

Ověřování modelu

Toto ověřování má potvrdit nebo zamítnout adekvátnost modelu. Můžeme mít podezření na neadekvátnost modelu, například z důvodu vysokých p-hodnot některých parametrů modelu. V tomto případě se doporučuje model přeparametrizovat a zkonstruovat znovu. Pokud se p-hodnoty parametrů sníží nebo výrazně klesne MSE nebo se residua ukáží korelovaná, pak původní model nahradíme novým. Na ověření nekorelovanosti residuí můžeme použít Portmanteau test. Konečné schválení modelu je v rámci posouzení statistika.

3.3.2 Specifikace modelu

Program vyžaduje zadat následující proměnné:

Time Series Variable - zde vybereme řadu, na kterou chceme model aplikovat.
Use Logarithms - zda chceme řadu před zpracováním logaritmovat (jedná se o logaritmus o základu 10).
Remove Mean - zda chceme řadu před zpracováním centrovat.
Remove Trend - metodou nejmenších čtverců se nalezne trend, který se od řady odečte. Místo tohoto se dá použít diferencování řady.
Regular AR - specifikuje nejvyšší počet parametrů autoregresního modelu.
Regular Diff. - specifikuje řád diference (0,1,2).
Regular MA - specifikuje nejvyšší počet parametrů procesu klouzavých součtů.
Seasonal AR - specifikuje nejvyšší počet parametrů sezónního autoregresního modelu.
Seasonal Diff. - specifikuje řád sezónní diference.
Seasonal MA - specifikuje nejvyšší počet parametrů sezónního procesu klouzavých součtů.
Seasons - počet sezón během roku.

First Season - pořadí 1. sezóny. Slouží pouze ke srozumitelnějšímu výpisu dat a grafů.
First Year - číslo 1. roku. Slouží také pouze ke srozumitelnějšímu výpisu dat a grafů.
Max Iterations - ne pro každou řadu iterace odhadu parametrů konverguje. Toto tedy určuje maximální počet iterací.
Convergence - při iteraci se vždy spočítá součet čtverců residuí. Když se při některém kroku toto změní o méně než tento parametr *Convergence*, iterace skončí. Tedy zvýšením tohoto parametru se sníží počet iterací a naopak.
Lambda - tento parametr ovlivňuje Marquartovu nelineární iterační proceduru. Většinou se ponechává defaultně nastavený na hodnotu 0.1.

Program po spuštění vypisuje (viz. Kapitola 4):

Minimization Phase Section - zde vypisuje hodnoty parametrů λ, φ, θ a MSE tak, jak se mění v průběhu iterační procedury.
Model Description Section - zde mimo jiné *Pseudo R-Squared*, což je statistika:

$$R^2 = 100(1 - \frac{SSE}{SST}), \tag{3.11}$$

kde SSE je součet čtverců residuí a SST je celkový součet čtverců pozorované řady po odečtení průměru. Dále hodnotu *SSE*, což je právě hodnota, kterou se iterační proces snaží minimalizovat. Dále *MSE* - ta byla popsána výše a *RSE*, což je odmocnina z *MSE*. Její výhodou je, že je ve stejných jednotkách jako původní časová řada.
Model Estimation Section - zde např. *T-Value*. Je to hodnota t-testu, který testuje nulovost daného parametru. Má tedy t rozdělení s $n-$(počet paramerů a diferencí modelu) stupni volnosti.
Asymptotic Correlation Matrix of Parameters - matice vzájemných korelací parametrů modelu.
Portmanteau Test - testuje nekorelovanost residuí do korelace ρ_k, testová statistika má χ^2 rozdělení o $K-p-q-P-Q$ stupních volnosti a je následující:

$$Q(k) = n(n+2)\sum_{j=1}^{k} \frac{r_j^2}{N-j}. \tag{3.12}$$

Kapitola 4

Aplikace modelů na reálná data

4.1 Data 1

Mějme měsíční měření časové řady. Na tato data budeme postupně v programu NCSS aplikovat modely popsané v předešlých kapitolách. Nejprve provedeme sezónní dekompozici.

Tabulka 4.1: Zaměstnanost mužů ve věku 16 - 19 let v USA v letech 1971 - 1981 (v tisících). Data byla převzata z literatury [4].

rok	led	úno	bře	dub	kvě	čer	červ	srp	zář	říj	lis	pro
1971	707	655	638	574	552	980	926	680	597	637	660	704
1972	758	835	747	617	554	929	815	702	640	588	669	675
1973	610	651	605	592	527	898	839	614	594	576	672	651
1974	714	715	672	588	567	1057	949	683	771	708	824	835
1975	980	969	931	892	828	1350	1218	977	863	838	866	877
1976	1007	951	906	911	812	1172	1101	900	841	853	922	886
1977	896	936	902	765	735	1234	1052	868	798	751	820	725
1978	821	895	851	734	636	994	990	750	727	754	792	817
1979	856	886	833	733	675	1004	956	777	761	709	777	771
1980	840	847	774	720	848	1240	1168	936	853	910	953	874
1981	1026	1030	946	860	856	1190	1038	883	843	857	1016	1003

4.1.1 Sezónní dekompozice

Forecast Summary Section

Forecast	(Mean) x (Trend) x (Cycle) x (Season)
Variable	muzi
Number of Rows	132
Mean	826.4697
Pseudo R-Squared	0.9311775
Forecast Std. Error	42.4505
Trend Equation	Trend = (0.835644) + (0.002479) * (Time Season Number)
Number of Seasons	12
First Year	1
First Season	1

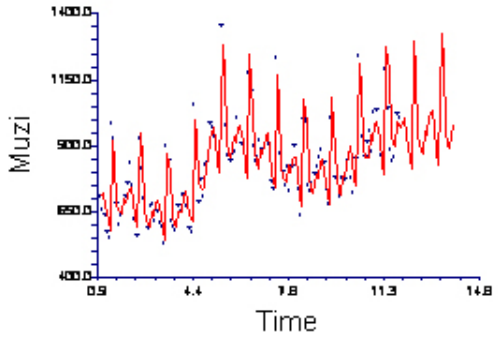
Z kolonky *Forecast* je zřejmé, že rozklad bude multiplikativní. Data, na která bude model aplikován, mají název *muzi* a jejich počet je 132. Dále je uveden mimo jiné *průměr*, rovnice *trendové přímky*, *počet sezón* a informace o prvním údaji, jako *rok* a jeho poloha v sezóně (roku). *Time Season Number* je pořadí naměřené hodnoty.

Seasonal Component Ratios

No.	**Ratio**	**No.**	**Ratio**	**No.**	**Ratio**	**No.**	**Ratio**
1	1.023273	2	1.041233	3	0.975670	4	0.881380
5	0.833239	6	1.329416	7	1.216548	8	0.957213
9	0.901875	10	0.884869	11	0.968265	12	0.951347

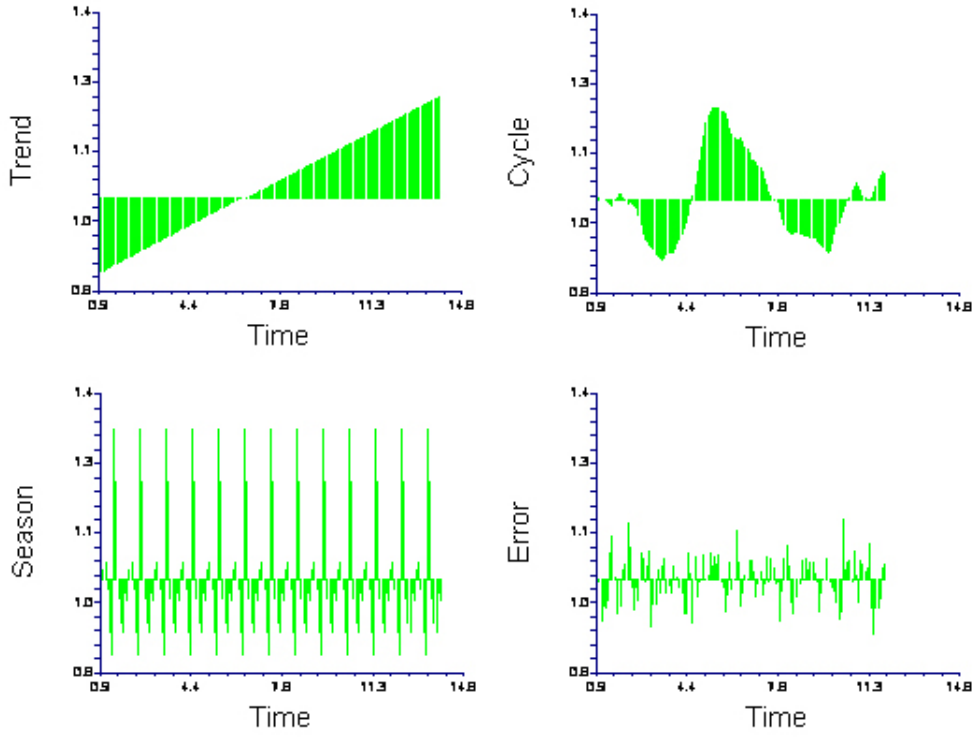
Toto jsou sezónní faktory, tedy hodnoty sezónní složky. Ta je periodická s periodou 12.

Forecast and Data Plot



Dále je uveden bodový graf dat s proloženou křivkou s předpověďmi. Vidíme, že sezónní charakter této řady je velmi významný.

Decomposition Ratio Plots



Zde vidíme grafy jednotlivých složek. Trend má vzestupný charakter, což odpovídá pohledu na naměřená data. Zajímavé je, že algoritmus rozpo-

znal na datech cyklickou složku. Sezónní složka má výraznou amplitudu a náhodná složka odpovídá bílému šumu.

Forecasts Section

Row No.	Year Season	Forecast Muzi	Actual Muzi	Residual	Trend Factor	Cycle Factor	Season Factor	Error Factor
133	12 1	985.4927			1.1653	1.0000	1.0233	1.0000
134	12 2	1004.923			1.1678	1.0000	1.0412	1.0000
135	12 3	943.6447			1.1702	1.0000	0.9757	1.0000
136	12 4	854.2546			1.1727	1.0000	0.8814	1.0000
137	12 5	809.3026			1.1752	1.0000	0.8332	1.0000
138	12 6	1293.949			1.1777	1.0000	1.3294	1.0000
139	12 7	1186.584			1.1802	1.0000	1.2165	1.0000
140	12 8	935.5972			1.1826	1.0000	0.9572	1.0000
141	12 9	883.3566			1.1851	1.0000	0.9019	1.0000
142	12 10	868.5125			1.1876	1.0000	0.8849	1.0000
143	12 11	952.35			1.1901	1.0000	0.9683	1.0000
144	12 12	937.6589			1.1926	1.0000	0.9513	1.0000

Vidíme, že u předpovědí byly náhodná a cyklická složka položeny 1. Sezónní složka se každý rok periodicky opakuje.

4.1.2 Exponenciální vyrovnávání

Aditivní

Nyní na shodná data provedeme exponenciální vyrovnávání s aditivní sezónní složkou:

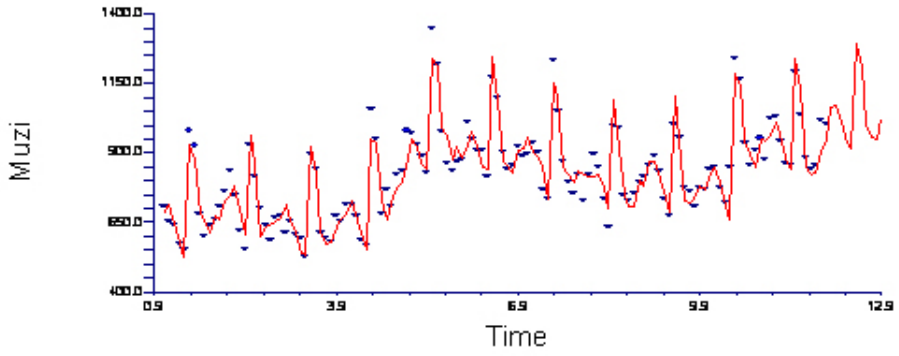
Forecast Summary Section

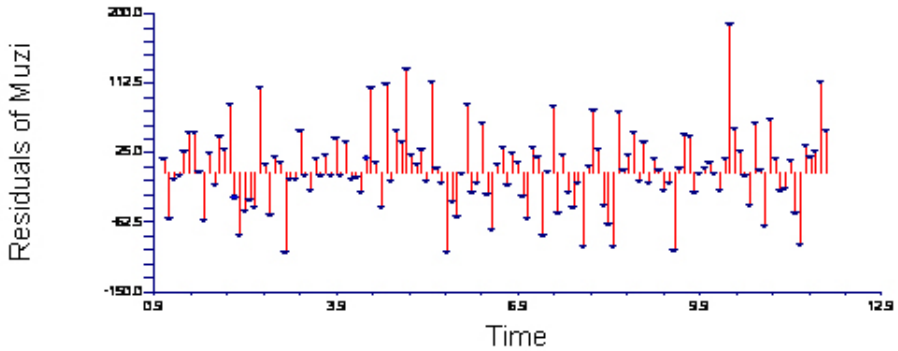
Mean Square Error	2521.409
Mean |Error|	37.75486
Mean |Percent Error|	4.555256
Forecast Method	Winter's with additive seasonal adjustment.
Search Iterations	129
Search Criterion	Mean Square Error
Alpha	0.6118585
Beta	6.31264E-06
Gamma	0.3811107
Intercept (A)	1297.669
Slope (B)	2.207226E-03

Season 1 Factor	-234.7966
Season 2 Factor	-230.0751
Season 3 Factor	-290.1971
Season 4 Factor	-362.7981
Season 5 Factor	-383.596
Season 6 Factor	-2.075533
Season 7 Factor	-88.27718
Season 8 Factor	-292.5083
Season 9 Factor	-341.7579
Season 10 Factor	-345.7422
Season 11 Factor	-277.6635
Season 12 Factor	-307.4202

Vidíme, že bylo provedeno exponenciální vyrovnávání Holt-Wintersovým algoritmem s aditivním sezónním nastavením. Řešení příslušných rovnic probíhá iteračně, je zde tedy uveden i počet iterací a iterační kritérium, které je nastaveno na minimalizaci MSE. α, β, γ jsou konstanty vyrovnávání, jsou určeny iterací a leží mezi 0 a 1. Parametry A a B jsou posunutí a směrnice trendové přímky pro fázi předpovídání do budoucna. V druhé kapitole, která se zabývala popisem výpočtu v NCSS, byly označeny jako a_t a b_t, tedy klademe $a_t = A$ a $b_t = B$. Předpověď lineárního trendu v čase k je tedy $A + B \cdot k$

Forecast and Residuals Plots





Grafy zakreslují data, proloženou křivku, předpovědi a residua.

Forecasts Section

Row No.	Date	Forecast Muzi
133	12 1	1063.167
134	12 2	1067.89
135	12 3	1007.77
136	12 4	935.1716
137	12 5	914.3759
138	12 6	1295.899
139	12 7	1209.699
140	12 8	1005.47
141	12 9	956.2228
142	12 10	952.2407
143	12 11	1020.322
144	12 12	990.5671

Toto jsou předpovídané hodnoty příštích 12 měsíců.

Multiplikativní

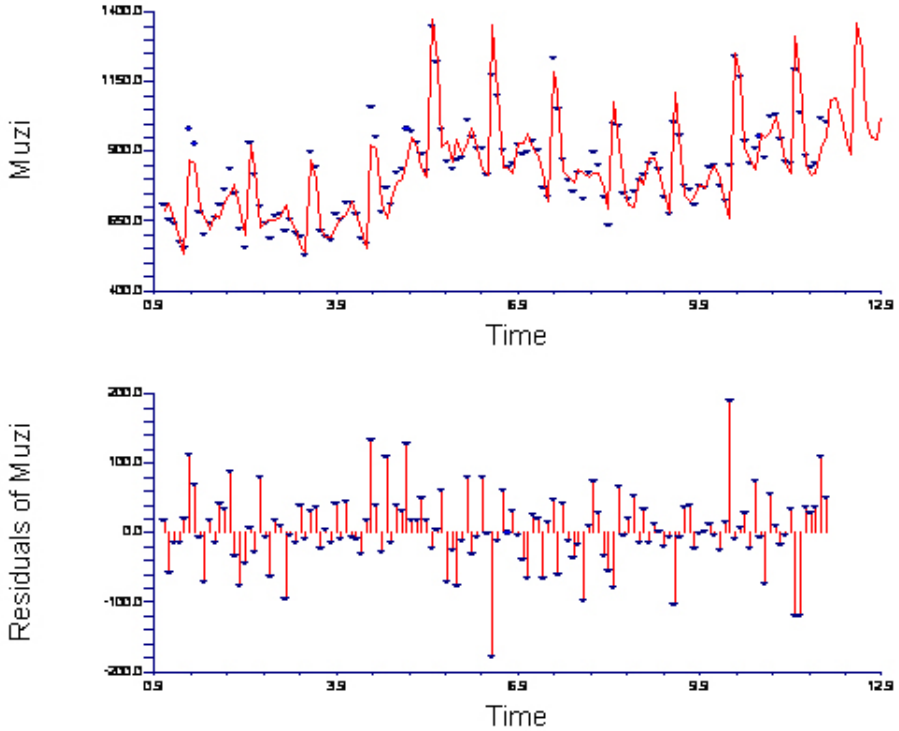
Exponenciální vyrovnávání s multiplikativní sezónní složkou:

Forecast Summary Section

Mean Square Error	2805.03
Mean |Error|	38.84326
Mean |Percent Error|	4.664131
Forecast Method	Winter's with multiplicative seasonal adjustment.
Search Iterations	129
Search Criterion	Mean Square Error
Alpha	0.6118585
Beta	6.31264E-06
Gamma	0.3811107
Intercept (A)	1045.968
Slope (B)	2.285455E-03
Season 1 Factor	1.033585
Season 2 Factor	1.043541
Season 3 Factor	0.9755233
Season 4 Factor	0.8869253
Season 5 Factor	0.8513904
Season 6 Factor	1.297284
Season 7 Factor	1.212406
Season 8 Factor	0.9694732
Season 9 Factor	0.9088718
Season 10 Factor	0.8979287
Season 11 Factor	0.9756466
Season 12 Factor	0.9474247

V tomto případě se jedná o Holt-Wintersův algoritmus s multiplikativním sezónním nastavením. Iterací je stejný počet jako v aditivním případě. Povšimněme si, že konstanty vyrovnávání α, β, γ byly určeny totožně.

Forecast and Residuals Plots



Graf dat a předpovědí vypadá velmi podobně jako graf v aditivním případě. Graf residuí má více vzdálených hodnot než v aditivním případě a tím se zvětšila hodnota MSE.

Forecasts Section

Row No.	Date	Forecast Muzi
133	12 1	1081.411
134	12 2	1091.831
135	12 3	1020.667
136	12 4	927.9714
137	12 5	890.794
138	12 6	1357.327
139	12 7	1268.523
140	12 8	1014.348
141	12 9	950.944
142	12 10	939.4963
143	12 11	1020.814
144	12 12	991.2879

Hodnoty předpovědí příštích 12 měsíců nejsou zásadně odlišné od hodnot exponenciálního vyrovnávání s aditivní sezónní složkou. Větší vychýlení hodnot by se mohlo ukázat u předpovědí dále do budoucna.

4.1.3 Sezónní ARIMA

Minimization Phase Section

V této fázi probíhá stanovení parametrů $\lambda, \varphi, \theta, \Phi, \Theta$ a MSE tak, jak se mění v průběhu iterační procedury.

Model Description Section

Series	LOG10(Muzi)
Model	Regular(1,0,3) Seasonal(7,2,1) Seasons = 12
Observations	132
Iterations	3
Pseudo R-Squared	87.179861
Residual Sum of Squares	0.1205544
Mean Square Error	1.255775E-03
Root Mean Square	3.543691E-02

Byl použit model $SARIMA(1,0,3) \times (7,2,1)_{12}$. Použití modelu s takto vysokým počtem parametrů není příliš časté, ale z důvodu splnění předpokladu nekorelovanosti residuí nebylo možno provést jinou volbu. Vhodnost použití tohoto modelu potvrdila i nízká hodnota MSE. Řada byla tedy nejdříve logaritmována a poté dvojnásobně sezónně diferencována. Kolonka *Iterations* udává počet iterací při určování hodnot $\lambda, \varphi, \theta, \Phi, \Theta$ a *Observations* počet naměřených dat časové řady. Časová řada nebyla před modelováním centrována. Délka sezóny je 12. Model má tedy tvar:

$$(1-\varphi_1 B)(1-B^{12}-B^{24})(1-\Phi_1 B^{12}-...-\Phi_7 B^{84})z_t =$$
$$= (1+\theta_1 B+\theta_2 B^2+\theta_3 B^3)(1+\Theta_1 B^{12})\varepsilon_t,$$

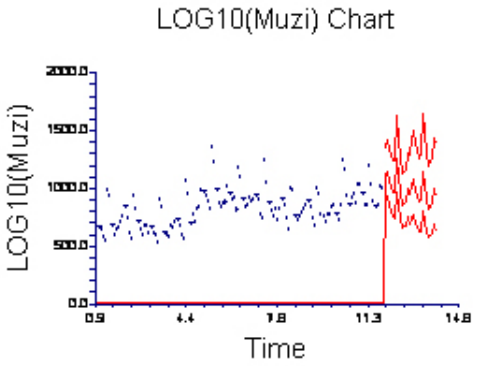
kde $z_t = log(y_t)$.

Model Estimation Section

Parameter Name	**Parameter Estimate**	**Standard Error**	**T-Value**	**Prob Level**
AR(1)	0.8073057	0.2073547	3.8934	0.000099
SAR(1)	-0.524031	0.1069282	-4.9008	0.000001
SAR(2)	-0.4684805	0.1054892	-4.4410	0.000009
SAR(3)	-0.3083988	0.1122569	-2.7473	0.006010
SAR(4)	-0.2602053	0.1110579	-2.3430	0.019131
SAR(5)	-4.626586E-02	0.111925	-0.4134	0.679339
SAR(6)	7.804021E-02	9.274586E-02	0.8414	0.400101
SAR(7)	0.2760307	7.973912E-02	3.4617	0.000537
MA(1)	0.1056408	0.2246244	0.4703	0.638141
MA(2)	0.0696576	0.1763885	0.3949	0.692909
MA(3)	-0.2213563	0.1118747	-1.9786	0.047860
SMA(1)	0.6371343	8.537184E-02	7.4630	0.000000

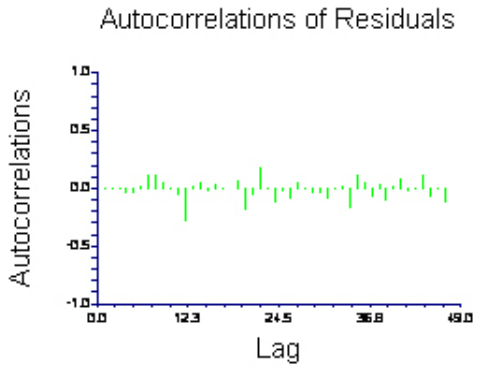
P-hodnoty hypotéz o nulovosti parametrů mají některé velice vysoké hodnoty, tedy nemůžeme jejich nulovost na 5 % hladině významnosti zamítnout. Mohli bychom pravděpodobně z modelu tyto parametry vyřadit, ale přechod k modelu nižších řádů není možný bez vyřazení parametrů SAR(7) nebo MA(3), jejichž p-hodnoty jsou dostatečně nízké. Ponecháváme tedy všechny parametry. NCSS vypisuje dále hodnoty testových statistik (*T-Value*), odhad hodnoty parametru (*Parameter Estimate*) a směrodatnou odchylku odhadu parametru od skutečné hodnoty(*Standard Error*).

Forecast and Data Plot



Graf dat a předpovědí. Všimněme si, že stochasticita modelu nám umožnila určit kromě bodových předpovědí také oblast 95 procentní pravděpodobnosti vývoje.

Autocorrelation Plot Section



Graf autokorelační funkce residuí. NCSS hodnoty residuí vypisuje spolu s hodnotou, jejímž překročením residua svědčí o své korelovanosti. Tato hodnota je zde 0.174078 a byla překročena jen ojediněle, tedy residua jsou téměř nekorelovaná. Podařilo se nám volbou vhodného modelu tento důležitý předpoklad splnit. Dále by program vypsal výsledek Portmanteau testu, není zde ale vzhledem ke své rozsáhlosti uveden.

4.2 Data 2

Tabulka 4.2: Nezaměstnanost žen ve věku 16 - 19 let v USA v letech 1961 - 1985 (v tisících). Data převzata z literatury [4].

rok	led	úno	bře	dub	kvě	čer	červ	srp	zář	říj	lis	pro
1961	375	384	383	326	344	375	419	424	429	399	376	288
1962	360	376	360	381	354	301	333	339	316	352	378	360
1963	388	398	377	383	449	415	429	369	414	462	447	403
1964	409	390	380	438	431	426	348	394	396	451	384	491
1965	466	454	442	475	401	406	385	380	422	397	430	433
1966	421	374	401	451	465	456	469	466	412	427	414	384
1967	328	395	381	360	383	383	403	425	422	414	382	390
1968	320	412	437	421	450	442	450	412	422	372	375	392
1969	356	392	426	442	426	406	392	426	445	464	379	409
1970	497	459	513	549	447	445	432	514	565	557	601	582
1971	587	560	590	556	582	527	585	556	574	556	582	583
1972	644	620	618	623	546	568	595	605	598	592	558	595
1973	549	637	568	605	594	567	545	545	592	576	593	603
1974	631	614	617	546	632	673	732	593	693	730	731	733
1975	802	755	805	751	855	769	800	825	799	802	765	827
1976	760	781	769	766	752	751	761	873	750	758	772	791
1977	813	781	797	802	782	838	756	764	796	781	780	679
1978	748	759	749	756	802	754	792	772	769	731	746	741
1979	712	723	698	746	754	735	722	737	728	773	723	741
1980	738	765	748	707	808	746	773	751	721	731	735	701
1981	762	783	796	803	806	765	781	768	812	854	858	818
1982	856	897	817	872	895	825	922	915	902	908	911	919
1983	861	827	855	867	836	916	828	835	792	771	757	756
1984	712	733	746	728	707	666	636	676	696	654	613	677
1985	705	680	699	650	687	638	670	555	631	676	659	689

4.2.1 Sezónní dekompozice

Forecast Summary Section

Forecast	(Mean) x (Trend) x (Cycle) x (Season)
Variable	zeny
Number of Rows	300
Mean	599.3367
Pseudo R-Squared	0.9670612
Forecast Std. Error	31.45931
Trend Equation	Trend = (0.567215) + (0.002915) * (Time Season Number)
Number of Seasons	12
First Year	1
First Season	1

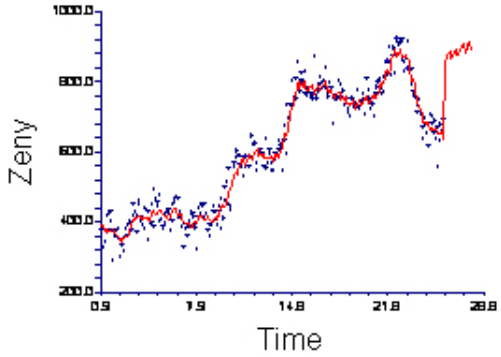
Nejprve program vypíše základní informace o časové řadě. Jedná se o multiplikativní sezónní dekompozici aplikovanou na data řady "zeny". Těchto měření je 300, dále je mimo jiné uveden průměr, rovnice přímky trendu, délka sezóny. *First Year* slouží pouze pro správný výpis dat a předpovědí, *First Season* použijeme v případě, že naměřená data začínají během roku.

Seasonal Component Ratios

No.	**Ratio**	**No.**	**Ratio**	**No.**	**Ratio**	**No.**	**Ratio**
1	0.989434	2	1.003301	3	1.006128	4	1.010074
5	1.016604	6	0.987282	7	0.996155	8	0.992654
9	1.005466	10	1.010242	11	0.989456	12	0.988193

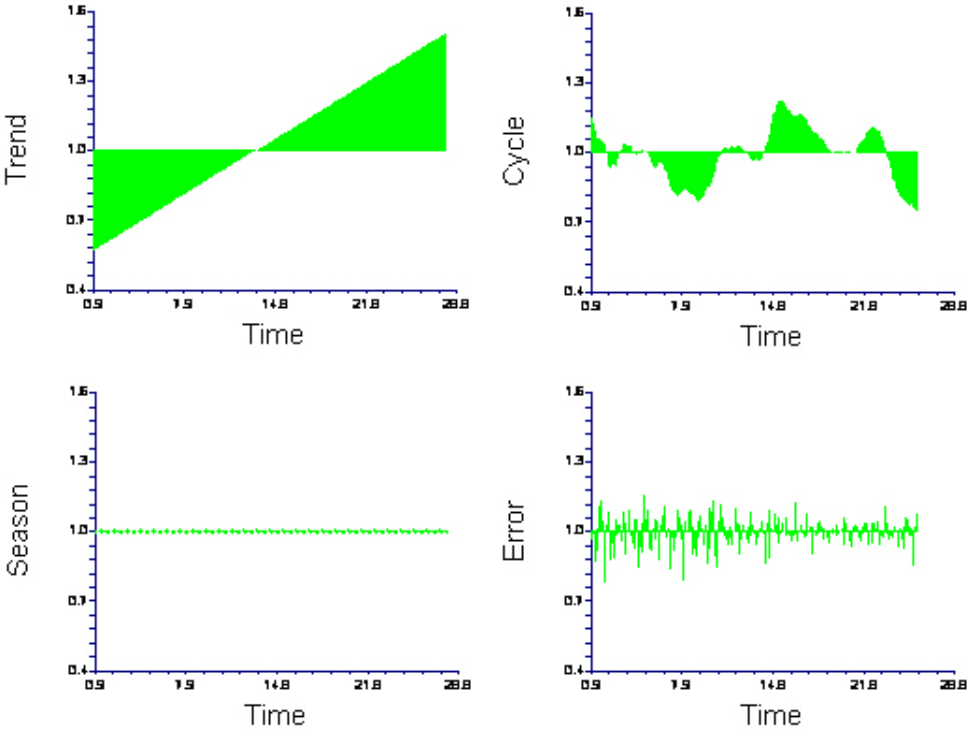
Toto jsou jednotlivé hodnoty sezónní složky. Periodicky se opakují. Vidíme, že jsou všechny velice blízké 1, tedy sezónní chování řady není příliš významné.

Forecast and Data Plot



Zde je graf naměřených dat, křivky dekompozice a 24 předpovědí. Jedná se o deterministický model, tudíž jsou předpovězeny pevné hodnoty.

Decomposition Ratio Plots



Toto jsou grafy všech čtyř složek. Průměr byl uveden v první části. Náhodná složka projevuje charakter bílého šumu. Časová řada nevykazuje

sezónní charakter, zato algoritmus odhalil cyklickou složku. Tu ale tento model neumí předpovídat do budoucna a tedy ve výpočtu předpovědí nebude zohledněna.

Forecast Section

Row No.	Year Season	Forecast zeny	Actual zeny	Residual	Trend Factor	Cycle Factor	Season Factor	Error Factor
313	1987 1	877.406			1.4796	1.0000	0.9894	1.0000
314	1987 2	891.4561			1.4825	1.0000	1.0033	1.0000
315	1987 3	895.7252			1.4854	1.0000	1.0061	1.0000
316	1987 4	901.0031			1.4883	1.0000	1.0101	1.0000
317	1987 5	908.6042			1.4913	1.0000	1.0166	1.0000
318	1987 6	884.122			1.4942	1.0000	0.9873	1.0000
319	1987 7	893.8078			1.4971	1.0000	0.9962	1.0000
320	1987 8	892.4012			1.5000	1.0000	0.9927	1.0000
321	1987 9	905.6756			1.5029	1.0000	1.0055	1.0000
322	1987 10	911.7429			1.5058	1.0000	1.0102	1.0000
323	1987 11	894.7117			1.5087	1.0000	0.9895	1.0000
324	1987 12	895.2962			1.5117	1.0000	0.9882	1.0000

Zde by program NCSS vypsal pro každý časový okamžik informaci o naměřené hodnotě, hodnotě trendu, hodnotě cyklické složky, hodnotě sezónní složky a hodnotě náhodné složky. Dále tyto hodnoty pro 24 předpovědí. Z důvodu rozsáhlosti této části jsou uvedeny pouze hodnoty posledních 12 předpovědí. *Cycle Factor* a *Error Factor* jsou pro předpovědi ponechány 1, jak bylo uvedeno v kapitole 2. Sloupec *Actual zeny* zůstal prázdný, jelikož se jedná o předpovědi a tedy naměřená data nejsou k dispozici. Sloupec *Residual* zůstal prázdný (nulový), protože multiplikativní faktor náhodné složky je u předpovědí 1 a hodnota residuí je tedy 0.

4.2.2 Exponenciální vyrovnávání

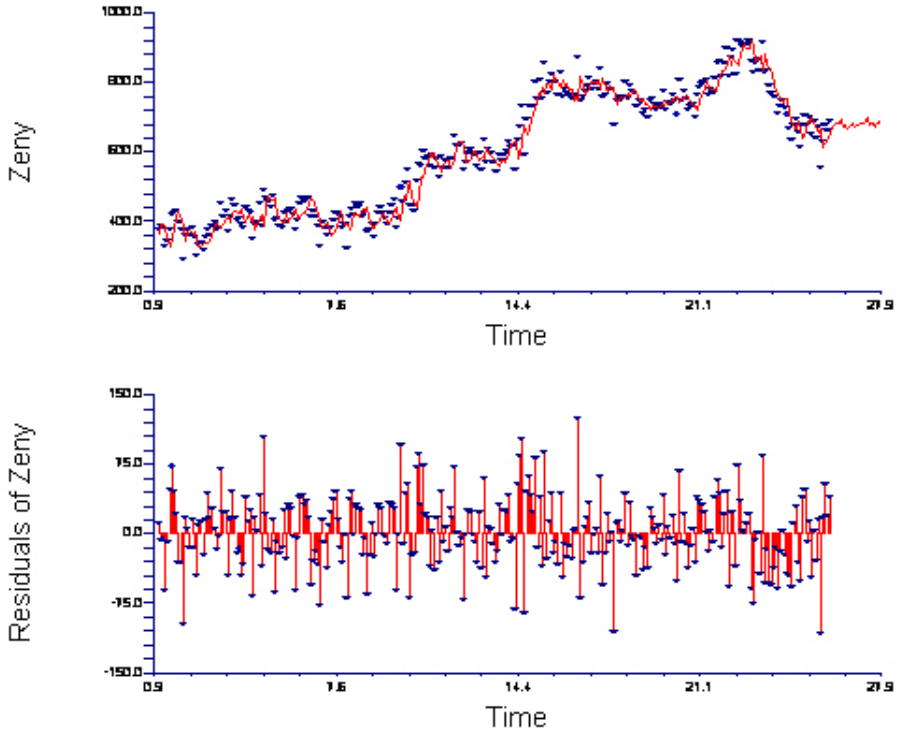
Aditivní

Forecast Summary Section

Variable	zeny
Number of Rows	300
Mean	599.3367
Pseudo R-Squared	0.951488
Mean Square Error	1457.61
Mean |Error|	30.00134
Mean |Percent Error|	5.47161
Forecast Method	Winter's with additive seasonal adjustment.
Search Iterations	137
Search Criterion	Mean Square Error
Alpha	0.4535435
Beta	1.490109E-03
Gamma	0.1327852
Intercept (A)	841.986
Slope (B)	0.3864124
Season 1 Factor	-286.7765
Season 2 Factor	-280.7172
Season 3 Factor	-278.99
Season 4 Factor	-278.4925
Season 5 Factor	-267.412
Season 6 Factor	-288.5301
Season 7 Factor	-284.6842
Season 8 Factor	-291.8466
Season 9 Factor	-283.8687
Season 10 Factor	-280.1209
Season 11 Factor	-290.4185
Season 12 Factor	-287.0593

Tato část shrnuje základní poznatky o řadě. Byl použit Holt-Wintersův algoritmus exponenciálního vyrovnávání s aditivním sezónním nastavením. Rozsah dat je 300, dále je uveden průměr, hodnoty *Pseudo R-Squared, MSE, MAE, MAPE* (jejich význam byl vysvětlen v kapitole 2). Parametry α, β, γ byly určeny interně programem pomocí *efficient grid-searching* algoritmu. A, B jsou posunutí a směrnice trendové přímky, která se použije při výpočtu předpovědí.

Forecast and Residuals Plots



Toto jsou grafy dat, exponenciálně vyrovnané křivky, předpovědí a příslušných residuí. Zde následuje ke každému časovému úseku výpis předpovězené hodnoty z minulé hodnoty, skutečně naměřené hodnoty a jejich rozdíl (residuum). Tato residua jsou zachycena ve zmíněném grafu residuí. Z důvodu rozsáhlosti je zde uvedeno pouze prvních 12 z celkových 24 předpovědí.

Forecasts Section

Row No.	Date	Forecast zeny	Actual zeny	Residuals
313	27 1	676.1566		
314	27 2	682.6023		
315	27 3	684.7159		
316	27 4	685.5999		
317	27 5	697.0668		
318	27 6	676.3351		
319	27 7	680.5674		
320	27 8	673.7914		
321	27 9	682.1558		
322	27 10	686.2899		
323	27 11	676.3787		
324	27 12	680.1244		

Poslední dva sloupce zůstaly prázdné, nemáme totiž k dispozici měření skutečných hodnot v časech předpovědí. Tedy ani residua, jakožto rozdíl těchto dvou hodnot nemají žádné hodnoty.

Multiplikativní

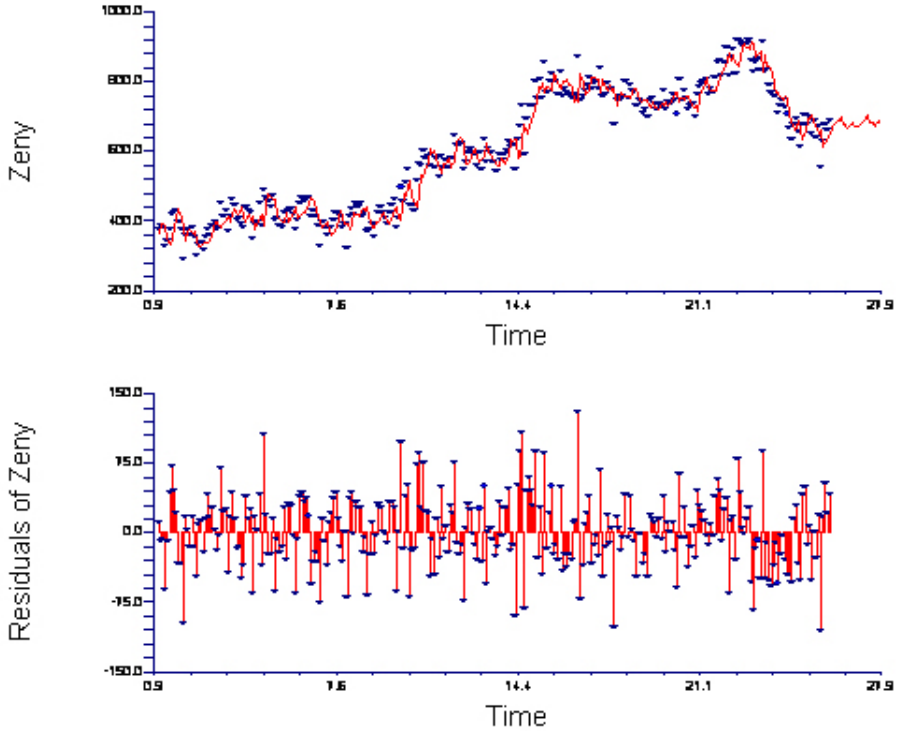
Forecast Summary Section

Variable	zeny
Number of Rows	300
Mean	599.3367
Pseudo R-Squared	0.949357
Mean Square Error	1521.64
Mean |Error|	30.90929
Mean |Percent Error|	5.607797

Forecast Method	Winter's with multiplicative seasonal adjustment.
Search Iterations	137
Search Criterion	Mean Square Error
Alpha	0.4535435
Beta	1.490109E-03
Gamma	0.1327852
Intercept (A)	560.3465
Slope (B)	0.387592
Season 1 Factor	0.9871608
Season 2 Factor	1.001904
Season 3 Factor	1.009796
Season 4 Factor	1.013632
Season 5 Factor	1.029862
Season 6 Factor	0.9958488
Season 7 Factor	0.998045
Season 8 Factor	0.9835187
Season 9 Factor	0.9985788
Season 10 Factor	1.004593
Season 11 Factor	0.9869116
Season 12 Factor	0.9901495

Tato část je analogická aditivnímu exponenciálnímu vyrovnávání, pouze s rozdílem, že se jedná o Holt-Wintersův algoritmus exponenciálního vyrovnávání s multiplikativním sezónním nastavením.

Forecast and Residuals Plots



Toto jsou grafy dat, vyrovnané křivky, předpovědí a residuí. Opět zde pro rozsáhlost uvedeme pouze prvních 12 z celkových 24 předpovědí časové řady.

Forecasts Section

Row No.	Date	Forecast zeny	Actual zeny	Residuals
313	27 1	672.9108		
314	27 2	683.3489		
315	27 3	689.1232		
316	27 4	692.1337		
317	27 5	703.6151		
318	27 6	680.7629		
319	27 7	682.6511		
320	27 8	673.0965		
321	27 9	683.7903		
322	27 10	688.2982		
323	27 11	676.5661		
324	27 12	679.1696		

4.2.3 Sezónní ARIMA

Minimization Phase Section

V této fázi probíhá stanovení parametrů $\lambda, \varphi, \theta, \Phi, \Theta$ a MSE tak, jak se mění v průběhu iterační procedury.

Model Description Section

Series	zeny-TREND
Model	Regular(2,0,1) Seasonal(2,0,1) Seasons = 12
Trend Equation	(335.9618)+(1.749999)x(date)
Observations	300
Iterations	12
Pseudo R-Squared	95.528054
Residual Sum of Squares	403096.4
Mean Square Error	1371.076
Root Mean Square	37.02805

Zde uvedeny informace o řadě jako název dat, rovnice trendové přímky, počet pozorování, počet iterací při zjišťování paramentrů $\lambda, \varphi, \theta, \Phi, \Theta$, dále MSE, MAE, MAPE, Pseudo R-Squared. Data nebyla diferencována, byl

pouze odečten trend. Zvolený model je tedy $SARIMA(2,0,1) \times (2,0,1)_{12}$ a rovnice modelu vypadá

$$(1 - \varphi_1 B - \varphi_2 B^2)(1 - \Phi_1 B^{12} - \Phi_2 B^{24}) z_t = (1 + \theta_1 B)(1 + \Theta_1 B^{12}) \varepsilon_t$$

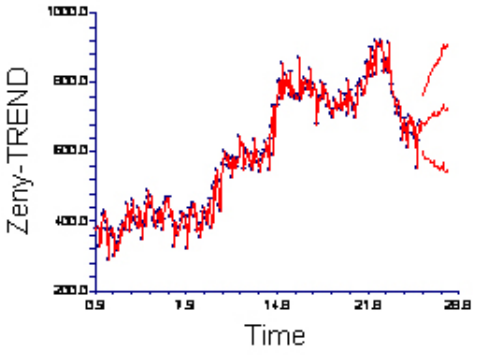
kde $z_t = y_t - Tr_t$.

Model Estimation Section

Parameter Name	**Parameter Estimate**	**Standard Error**	**T-Value**	**Prob Level**
AR(1)	1.060052	0.106486	9.9548	0.000000
AR(2)	-5.919621E-02	0.1065457	-0.5556	0.578488
SAR(1)	0.4324459	0.1944548	2.2239	0.026156
SAR(2)	-0.1645171	6.404044E-02	-2.5690	0.010201
MA(1)	0.5516425	8.971749E-02	6.1487	0.000000
SMA(1)	0.4749414	0.1969713	2.4112	0.015899

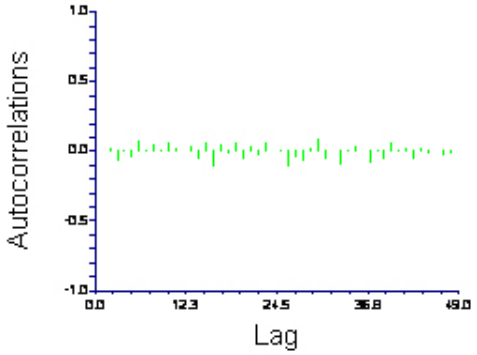
Zde vidíme spočítané hodnoty parametrů modelu, jejich testovou statistiku, směrodatnou odchylku a p-hodnotu. Parametr AR(2) má p-hodnotu 0.578488, tedy nezamítáme jeho nulovost na hladině do 57 procent. Navíc vzhledem k jeho malé hodnotě by bylo vhodné zjednodušit model na řád autoregrese 1.

Forecast and Data Plot



Toto je graf naměřených dat, křivky modelu a předpovědí. Úsek předpovědí zahrnuje bodové předpovědi a oblast s 95 procentní pravděpodobností vývoje v jednotlivých měsících.

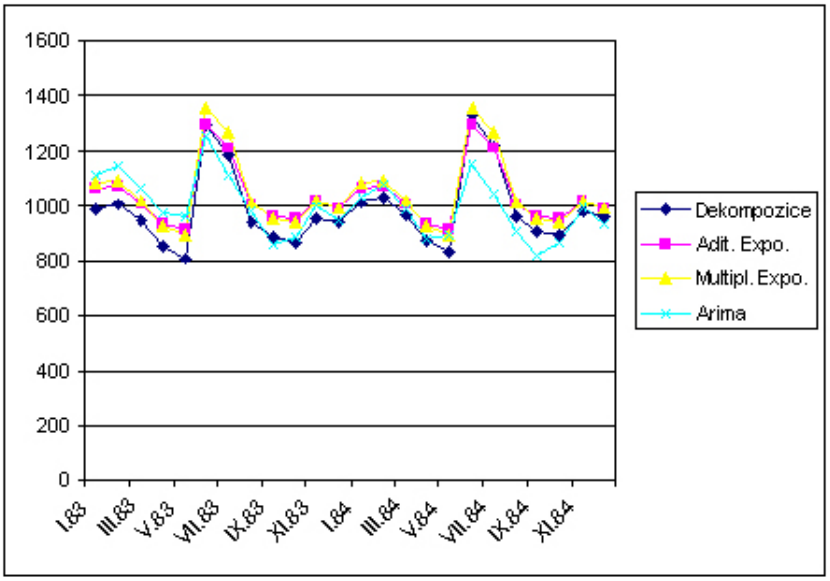
Autocorrelation Plot Section



Autokorelační funkce residuí. NCSS vypisuje tyto hodnoty do tabulky. Korelace je významná při překročení v absolutní hodnotě 0.115470. To se stalo pouze v jediném případě a o zanedbatelnou hodnotu. Můžeme tedy residua pokládat za nekorelovaná.

4.3 Shrnutí

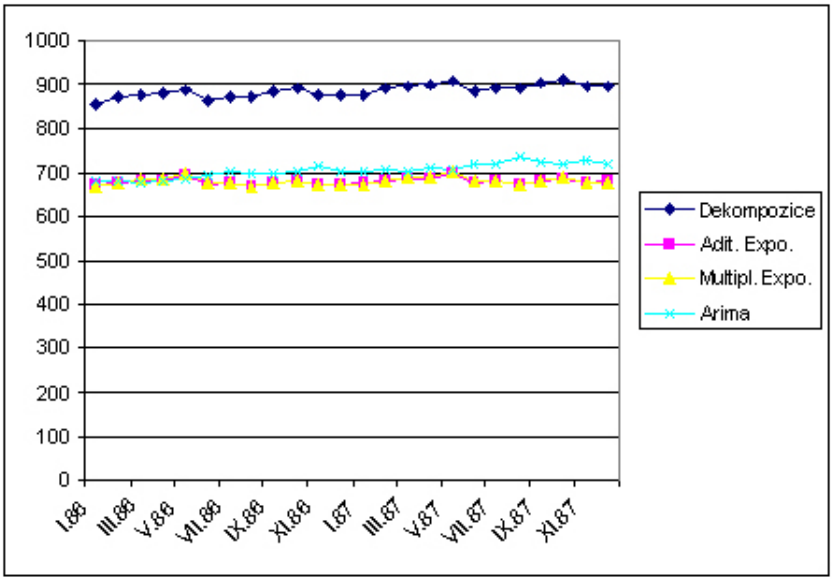
Data 1



Zde vidíme předpovědi 24 budoucích hodnot podle výše zmíněných me-

tod. Neopomenutelný je sezónní charakter předpovědí.

Data 2



Data této časové řady, tedy ani předpovědi nejeví tak výrazný sezónní charakter, jako data předchozí časové řady. Všimněme si, že předpovědi sezónní dekompozice vykazují daleko vyšší hodnoty než ostatní modely. To je způsobeno tím, že ke konci naměřených dat má cyklická složka výrazně záporné hodnoty. Při tvorbě předpovědí se ale položí rovna 1, tedy hodnoty předpovědí se v tomto případě zvýší.

Literatura

[1] RNDr. Tomáš Cipra, CSc.: *Analýza časových řad s aplikacemi v ekonomii*, Alfa, SNTL, Praha, 1986.

[2] Anděl, J: *Statistická analýza časových řad*, SNTL, Praha, 1976.

[3] Zuzana Prášková, Petr Lachout: *Základy náhodných procesů*, Nakladatelství Karolinum, Praha, 2005.

[4] William W. S. Wei: *Time Series Analysis*, Addison - Wesley, New York, 1990.