VYSOKÁ ŠKOLA EKONOMICKÁ V PRAZE
Fakulta informatiky a statistiky

ANALÝZA EKONOMICKÝCH ČASOVÝCH ŘAD S PŘÍKLADY

Josef Arlt
Markéta Arltová
Eva Rublíková

2002

Recenzenti:
Prof. Ing. František Fabian, CSc.
Doc. Ing. Jiří Trešl, CSc.

© Doc. Ing. Josef Arlt, CSc., Ing. Markéta Arltová, Ph.D., Doc. RNDr. Eva Rublíková, CSc. - Praha 2002

OBSAH

ÚVOD 5
1. ZÁKLADNÍ POJMY A GRAFICKÁ ANALÝZA 7
1.1 Ekonomické časové řady a jejich druhy 7
1.2 Grafická analýza časových řad 7
1.2.1 Spojnicový graf jedné časové řady 8
1.2.2 Spojnicový graf dvou a více časových řad 9
1.2.3 Krabičkový graf 10
1.2.4 Graf sezónních hodnot 11
1.2.5 Graf ročních hodnot sezónních časových řad 12
1.3 Základní charakteristiky časových řad 13
1.3.1 Popisné charakteristiky 13
1.3.2 Míry dynamiky 14
1.3.3 Cvičení 19
2. MODELOVÁNÍ TRENDU A SEZÓNNÍ SLOŽKY 20
2.1 Dekompozice časových řad 20
2.2 Základní metody modelování trendové složky, konstrukce předpovědí 21
2.2.1 Analýza trendu 21
2.2.2 Analýza trendu ve STATGRAPHICSu 31
2.2.3 Klouzavé průměry 43
2.2.4 Klouzavé průměry ve STATGRAPHICSu 46
2.2.5 Exponenciální vyrovnávání 51
2.2.6 Exponenciální vyrovnávání ve STATGRAPHICSu 56
2.2.7 Cvičení 61
2.3 Základní metody modelování sezónní složky, konstrukce předpovědí 62
2.3.1 Identifikace sezónnosti 62
2.3.2 Identifikace sezónnosti ve STATGRAPHICSu 63
2.3.3 Sezónní dekompozice, určení sezónních indexů 65
2.3.4 Sezónní dekompozice, určení sezónních indexů ve STATGRAPHICSu 66
2.3.5 Předpovídání časových řad po sezónní dekompozici 70
2.3.6 Regresní metoda modelování sezónnosti 71
2.3.7 Holtovo-Wintersovo exponenciální vyrovnávání 79
2.3.8 Holtovo-Wintersovo exponenciální vyrovnávání ve STATGRAPHICSu 81
2.3.9 Cvičení 83
3. BOXOVA-JENKINSOVA METODOLOGIE 85
3.1 Stochastický proces, stacionarita a autokorelační struktura 85
3.2 Stacionárních procesy 86
3.2.1 Procesy AR 86
3.2.2 Procesy MA 89
3.2.3 Procesy ARMA 92

3.3 Integrované procesy 94
3.3.1 Proces náhodné procházky 94
3.3.2 Procesy ARIMA 95
3.4 Sezónní procesy 95
3.4.1 Procesy SARIMA 95
3.5 Identifikace a ověřování modelu, konstrukce předpovědí 96
3.5.1 Identifikace modelu 96
3.5.2 Diagnostická kontrola modelu 97
3.5.3 Výpočet předpovědí 98
3.5.4 Boxova-Jenkinsova metodologie ve STATGRAPHICSu 99
3.5.5 Cvičení 130
PŘÍLOHY 131
PŘÍLOHA A – Použité časové řady 133
PŘÍLOHA B – Kritické hodnoty pro Darbinův-Watsonův test 141
PŘÍLOHA C – Příklady použití operátorů ve STATGRAPHICSu 143
LITERATURA 145

ÚVOD

Skripta „Analýza ekonomických časových řad s příklady“ jsou určena pro posluchače studijního programu „Statistické a pojistné inženýrství“ na VŠE v Praze a posluchačům ostatních programů, kteří si z volitelných kurzů na VŠE vybrali ty, jenž úzce souvisí s řešením praktických problémů pomocí analýzy časových řad. Předkládaná učební pomůcka předpokládá znalosti obsahu celoškolsky povinných kurzů „Statistika A“ (STP201), „Statistika B“ (STP202) resp. Statistika A+B (STP203). Obsahově může sloužit jako samostatný text, nebo jako doplněk skript „Úvod do analýzy ekonomických časových řad“ autorů J. Kozáka, R. Hindlse a J. Arlta z roku 1994.

Předkládaný text se skládá ze tří kapitol - Základní pojmy (autory jsou J. Arlt a M. Arltová), Modelování trendu a sezónní složky (E. Rublíková) a Boxova-Jenkinsova metodologie (J. Arlt a M. Arltová). V každé kapitole jsou uvedeny základní pojmy a je vysvětlena metodologie. Následuje řešení konkrétního problému spjatého s uvedenou metodou ve formě vzorového příkladu. Tyto příklady obsahují nejen tabulkové a grafické znázornění výsledků analýzy a jejich interpretaci, ale i popis postupu řešení ve statistickém paketu STATGRAPHICS Plus for Windows. V závěru kapitoly jsou vždy uvedeny neřešené příklady.

V příkladech jsou využity reálné časové řady, jejich označení vychází z principu, jakým jsou označena data uložená paketem STATGRAPHICS, jméno proměnné se skládá ze dvou částí oddělených podtržítkem (např. HDP_CR), první část označuje analyzovaný ukazatel a druhá část identifikuje zemi, jenž je zdrojem dat. Seznam použitých časových řad, jejich popis a hodnoty jsou uvedeny v příloze. V elektronické podobě jsou použité časové řady v souboru *Data.sf3* a *Data.xls* k dispozici na internetu na adresách

http://nb.vse.cz/~arltova/hlavni/vyuka/data/crsbirka02/data.sf3 (resp. .xls)

a http://nb.vse.cz/~arlt/hlavni/vyuka/data/crsbirka02/data.sf3 (resp. .xls).

Z důvodu využití statistického paketu STATGRAPHICS při řešení příkladů, doporučujeme jako vhodný doplněk skripta STATGRAPHICS for Windows (Zadávání úloh) autorů M. Arltové, M. Matušů a J. Kozáka.

Autoři

1. ZÁKLADNÍ POJMY

1.1 Ekonomické časové řady a jejich druhy

Jedním z důležitých úkolů statistických analýz ekonomických jevů je zkoumání jejich dynamiky. Empirická pozorování v ekonomické oblasti jsou často uspořádána do časové řady. Ekonomickou časovou řadou se rozumí řada hodnot jistého věcně a prostorově vymezeného ukazatele, která je uspořádána v čase směrem od minulosti do přítomnosti. Takto definovanou časovou řadu budeme zapisovat jako y_t, $t = 1, ..., T$.

Základní členění ekonomických časových řad lze ilustrovat následujícími příklady:

a) vývoz České republiky v letech 1991 až 1999,

b) počet zaměstnanců jisté firmy k prvnímu dni jednotlivých měsíců let 1991 až 2000,

c) produktivita práce v průmyslu České republiky v letech 1989 až 1998.

Časová řada ad a) je řadou intervalového ukazatele (*intervalová časová řada*). Velikost hodnoty tohoto ukazatele závisí na délce časového intervalu sledování. Typickými intervalovými ukazateli jsou extenzitní ukazatele, jejich příkladem může být objem výroby, spotřeba surovin atd. Časová řada ad b) se nazývá časovou řadou okamžikového ukazatele (*okamžiková časová řada*). Okamžikovým ukazatelem je ukazatel vztahující se k jistému okamžiku. Hodnota takového ukazatele nezávisí na délce časového intervalu sledování. Příkladem okamžikového ukazatele je počet pracovníků jisté firmy k určitému datu. Časová řada ad c) je řadou odvozené charakteristiky. Tento typ časových řad se získá z intervalových nebo okamžikových časových řad, např. časová řada produktivity práce se odvozuje jako podíl časové řady produkce a časové řady počtu pracovníků.

Ekonomické časové řady dále dělíme na dlouhodobé a krátkodobé. Hodnoty dlouhodobých časových řad jsou sledovány v ročních či delších časových úsecích. Hodnoty krátkodobých časových řad jsou sledovány v úsecích kratších než je jeden rok. Příkladem krátkodobých časových řad jsou časové řady čtvrtletní, měsíční, týdenní atd. Jak uvidíme dále, toto členění je důležité při zkoumání jednotlivých složek časových řad.

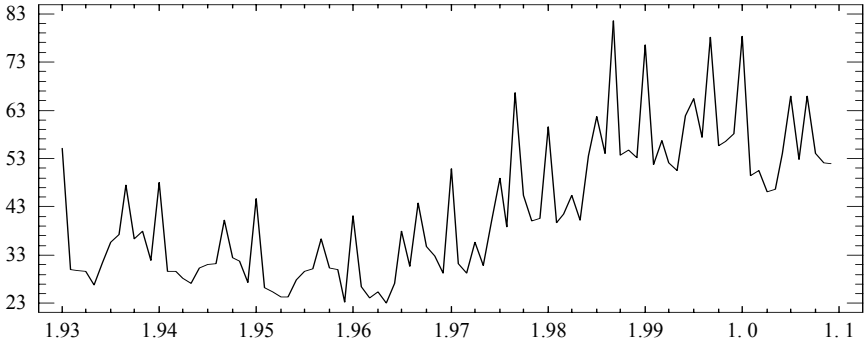
1.2 Grafická analýza časových řad

Jedním ze základních prostředků prezentace časových řad je jejich graf. Nejčastěji se graficky znázorňují původní hodnoty časové řady, nebo kumulativní časové řady, které vznikají postupným načítáním (kumulováním) jednotlivých hodnot (u okamžikových časových řad nemají smysl, neboť výše jejich hodnot nezávisí na daném časovém intervalu). Často se ale časové řady zobrazují tak, aby více vynikly jejich charakteristické vlastnosti a rysy. K tomu slouží speciální typy grafů.

1.2.1 Spojnicový graf jedné časové řady

Prvotní informace pro analýzu časových řad získáme ze spojnicových grafů. Jejich princip spočívá v zakreslení jednotlivých hodnot časové řady do souřadných os, na kterých jsou vyznačeny příslušné stupnice. Na osu horizontální se vynáší časová proměnná a na osu vertikální hodnoty časové řady nebo její funkce.

Příkladem může být obr. 1.1 zobrazující měsíční časovou řadu počtu nově registrovaných uchazečů o zaměstnání v České republice (NezNov_CR) v letech 1993 - 2000 v tis. osob.



Obr. 1.1: Spojnicový graf měsíční časové řady počtu nově registrovaných uchazečů o zaměstnání v ČR v období 1/1993 - 12/2000 (v tis. osob)

Pro zobrazení spojnicového grafu použijeme ve STATGRAPHICSu proceduru

Descriptive Methods

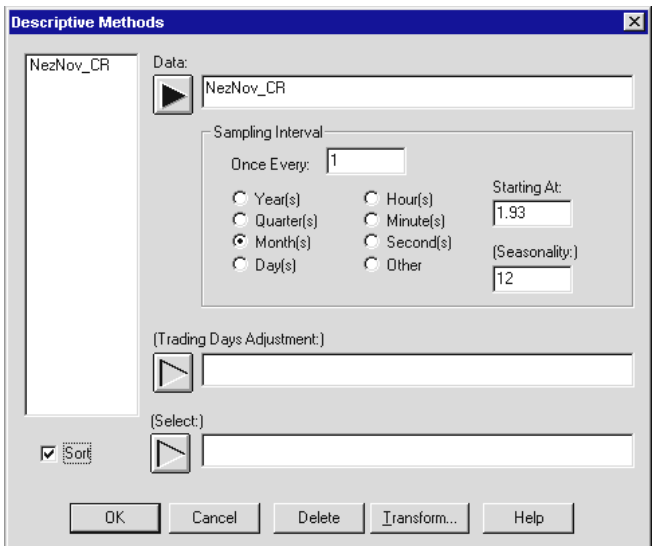
Special

⇒ **Time-Series Analysis**

⇒ **Descriptive Methods**

---

kde po vyplnění vstupního panelu (obr. 1.2) zvolíme Horizontal Time Sequence Plot z nabídky *Graphical Options*, dále už jen .



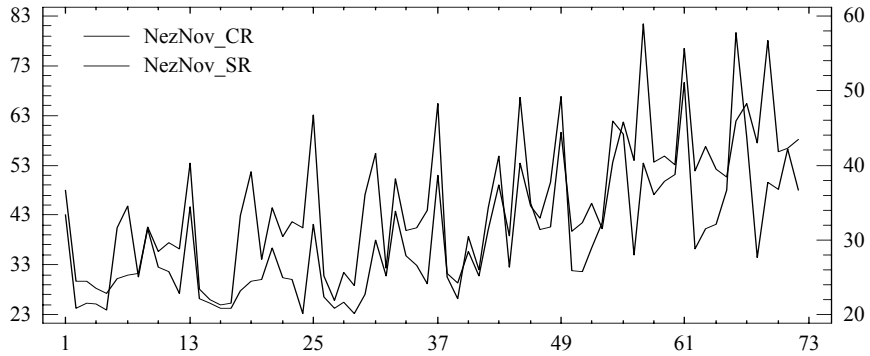
Obr. 1.2: Vyplnění vstupního panelu Descriptive Methods

U grafů časových řad je důležité správné nastavení časové osy, proto v části Sampling Interval vstupního panelu označíme v případě této měsíční časové řady položku Month(s) a do Starting At vepíšeme 1.93 (časová řada je měsíční od ledna 1993), dále vyplníme políčko Seasonality číslem 12, charakterizujícím počet sezón měsíčních časových řad, tj. počet hodnot časové řady získaných v kalendářním roce.

1.2.2 Spojnicový graf dvou a více časových řad

Do spojnicového grafu můžeme zakreslit i více časových řad. V případě, že zobrazujeme např. dvě časové řady lišící se měřítkem, je možné použít kromě levé i pravou vertikální osu.

Jako příklad lze zvolit grafické srovnání měsíčních časových řad počtu nově registrovaných uchazečů o zaměstnání v ČR (NezNov_CR) a SR (NezNov_SR) v letech 1994 - 1999 v tis. osob.



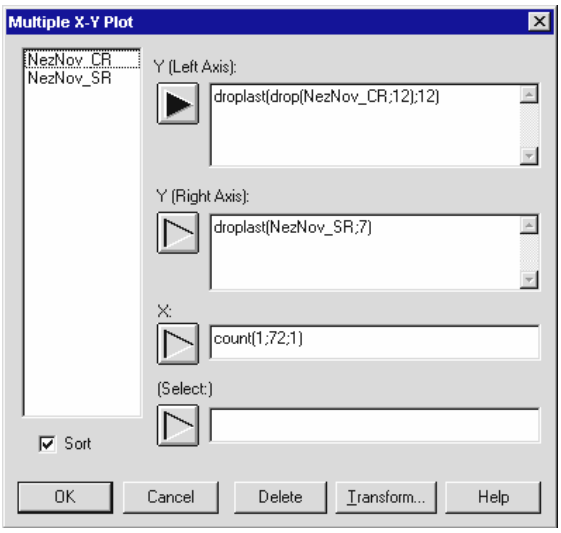
Obr. 1.3: Spojnicový graf měsíčních časových řad počtu nově registrovaných uchazečů o zaměstnání v ČR a ve SR v období 1/1994 - 12/1999 (v tis. osob)

Pro zobrazení tohoto typu grafu nejlépe vyhovuje procedura **Multiple X-Y Plot**

Plot

⇒ **Scatterplots**

⇒ **Multiple X-Y Plot**.



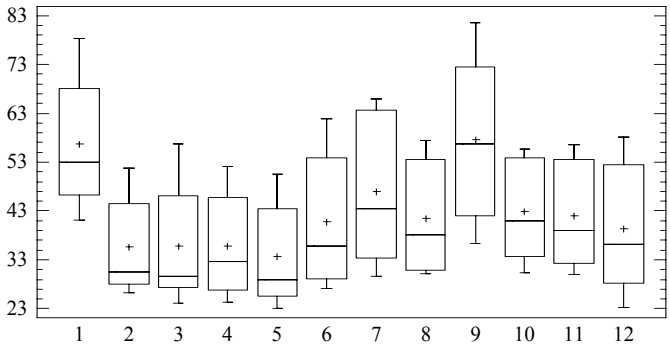
Obr. 1.4: Vyplnění vstupního panelu Multiple X-Y Plot

Vyplnění vstupního panelu je zřejmé z obr. 1.4. Abychom mohli graf zobrazit jako na obr. 1.3, je nutné nejprve zkrátit původní časové řady NezNov_CR a NezNov_SR ze souboru *Data.sf3* tak, aby obsahovaly údaje pouze za roky 1994 - 1999. Původní časová řada NezNov_CR je od ledna 1993 do prosince 2000 a řada NezNov_SR od ledna 1994 do července 2000. Pomocí operátorů je tedy třeba zkrátit NezNov_CR o 12 hodnot zepředu i zezadu a NezNov_SR o 7 hodnot zezadu. Ke zkrácení časové řady zepředu využijeme operátor `DROP` a pro zkrácení zezadu operátor `DROPLAST` (viz příklady použití operátorů v příloze). Do řádku X zadáváme časovou proměnnou, jíž je posloupnost hodnot od 1 do 72 vytvořená pomocí operátoru `COUNT`, nebo vypíšeme čísla z klávesnice do tabulkového editoru STATGRAPHICSu.

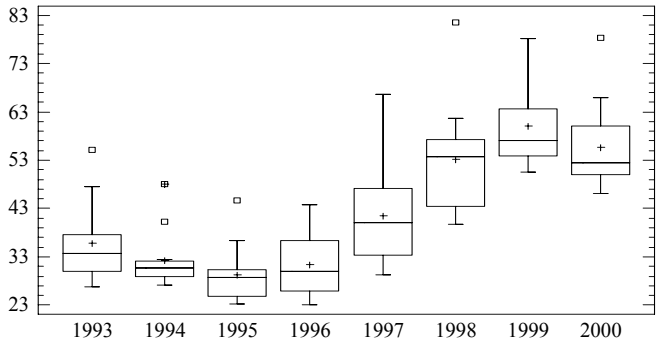
1.2.3 Krabičkový graf

V některých případech je užitečné provést detailnější pohled na časovou řadu. Krabičkový graf na rozdíl od jiných grafů obsahuje souhrnné charakteristiky zkoumané časové řady. Tento graf umožní odhalit některé důležité vlastnosti řady, které z jiných grafů nejsou zřetelné. Jeho základním prvkem je krabička, jejíž dolní a horní hrana je tvořena 25% a 75% kvartilem, uvnitř je vyznačen medián a symbolem „+“ aritmetický průměr. Na koncích svislých čar vycházejících z krabičky leží hodnoty minima a maxima. Protože délka této svislé čáry může být maximálně 1,5x delší než krabička, jsou hodnoty přesahující tyto hranice označovány jako odlehlé a jsou zakresleny jako samostatné body.

Jako příklad uvádíme krabičkový graf měsíční časové řady počtu nově registrovaných uchazečů o zaměstnání v ČR (NezNov_CR) v letech 1993 - 2000 v tis. osob zobrazený na obr. 1.5a podle měsíců a na obr. 1.5b podle roků.



Obr. 1.5a: Krabičkový graf měsíční časové řady počtu nově registrovaných uchazečů o zaměstnání v ČR v období 1/1993 - 12/2000 (v tis. osob) zobrazený podle měsíců



Obr. 1.5b: Krabičkový graf měsíční časové řady počtu nově registrovaných uchazečů o zaměstnání v ČR v období 1/1993 - 12/2000 (v tis. osob) zobrazený podle roků

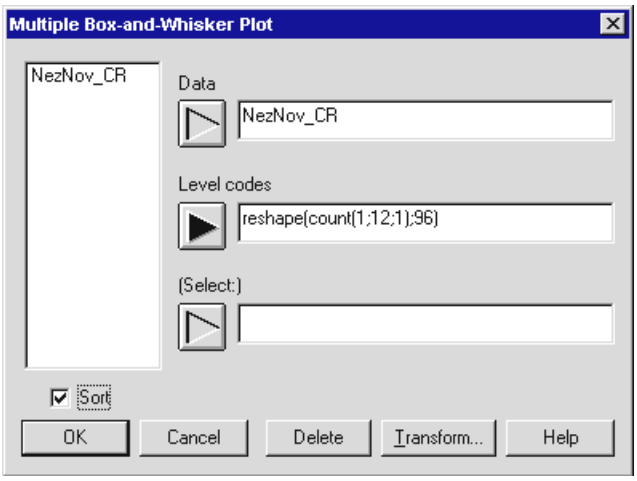
K zobrazení tohoto typu krabičkového grafu použijeme proceduru **Multiple Box-and-Whisker Plot**

Plot
⇒ **Exploratory Plots**
⇒ **Multiple Box-and-Whisker Plot**.

Do vstupního panelu vložíme do řádku Data časovou řadu NezNov_CR a do Level codes časovou proměnnou. Časová proměnná musí být v tomto případě modifikována tak, abychom dosáhli předem zvoleného zobrazení. Časová řada NezNov_CR je měsíční, můžeme ji tedy zobrazit podle měsíců (obr. 1.5a) nebo podle roků (obr. 1.5b).

Pro zobrazení podle měsíců bude časová proměnná obsahovat 8x se opakující posloupnost hodnot 1 - 12 (tj. 96 hodnot), kterou lze buď vypsat z klávesnice do tabulkového editoru STATGRAPHICSu, nebo je možné zapsat příkaz vytvořený pomocí operátorů `RESHAPE` a `COUNT` přímo do řádku Level codes (obr. 1.6)

```
reshape(count(1;12;1);96).
```



Obr. 1.6: Vyplnění vstupního panelu Multiple Box-and-Whisker Plot

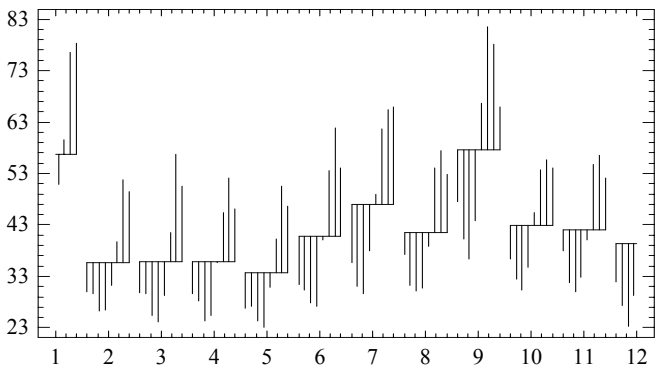
Chceme-li zobrazit časovou řadu podle let, bude časová proměnná obsahovat posloupnost hodnot 1993 - 2000, přičemž se každá hodnota za sebou opakuje 12x. Tuto posloupnost lze opět buď vypsat z klávesnice do tabulkového editoru STATGRAPHICSu, nebo lze použít příkaz vytvořený pomocí operátorů `REP` a `COUNT`

```
rep(count(1993;2000;1);12).
```

1.2.4 Graf sezónních hodnot

Tento graf se používá při analýze sezónních časových řad. Zobrazuje hodnoty časové řady uspořádané podle jednotlivých sezón. Vodorovné čáry představují průměrnou úroveň hodnot v jednotlivých sezónách za všechny roky časové řady. Svislé čáry znázorňují odchylky skutečných hodnot od průměru pro každou sezónu.

Příkladem může být graf měsíční časové řady počtu nově registrovaných uchazečů o zaměstnání v ČR (NezNov_CR) v letech 1993 - 2000 v tis. osob na obr. 1.7.



Obr. 1.7: Graf sezónních hodnot měsíční časové řady počtu nově registrovaných uchazečů o zaměstnání v ČR v období 1/1993 - 12/2000 (v tis. osob)

Graf sezónních hodnot zobrazíme v proceduře **Seasonal Decomposition**

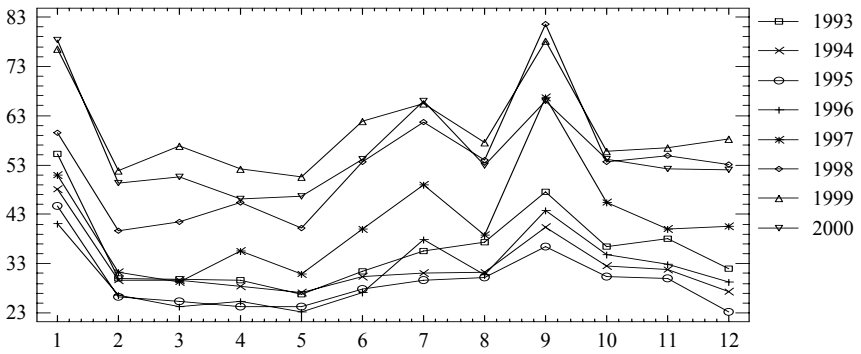
Special
⇒ **Time-Series Analysis**
⇒ **Seasonal Decomposition**

---

volbou Seasonal Subseries Plot z nabídky .

1.2.5 Graf ročních hodnot sezónních časových řad

Graf ročních hodnot sezónních časových řad zobrazuje hodnoty časové řady uspořádané podle roků a tak charakterizuje, jak se v jednotlivých letech liší úroveň hodnot v daných sezónách za celou časovou řadu.



Obr. 1.8: Graf ročních hodnot měsíční časové řady počtu nově registrovaných uchazečů o zaměstnání v ČR v období 1/1993 - 12/2000 (v tis. osob)

Graf ročních hodnot sezónních časových řad zobrazíme v proceduře **Seasonal Decomposition**, kde zvolíme Annual Subseries Plot z nabídky .

1.3 Základní charakteristiky časových řad

1.3.1 Popisné charakteristiky

Při práci s časovými řadami je někdy důležité zjistit jejich průměrné hodnoty. Průměrná hodnota intervalové časové řady se vypočítá pomocí *prostého aritmetického průměru*

$$\bar{y} = \frac{\sum_{t=1}^{T} y_t}{T}. \tag{1.1}$$

Průměrná hodnota okamžikové časové řady y_t, $t = 1, ..., T$ se počítá pomocí *chronologického průměru*. Při stejné vzdálenosti mezi jednotlivými okamžiky sledování se používá prostý chronologický průměr

$$\bar{y} = \frac{\frac{y_1 + y_2}{2} + \frac{y_2 + y_3}{2} + \ldots + \frac{y_{T-1} + y_T}{2}}{T-1} = \frac{\frac{1}{2} y_1 + \sum_{t=2}^{T-1} y_t + \frac{1}{2} y_T}{T-1}. \tag{1.2}$$

Při různé vzdálenosti jednotlivých okamžiků sledování se používá vážený chronologický průměr

$$\bar{y} = \frac{\frac{y_1 + y_2}{2} d_2 + \frac{y_2 + y_3}{2} d_3 + \ldots + \frac{y_{T-1} + y_T}{2} d_T}{d_2 + d_3 + \ldots + d_T}, \tag{1.3}$$

kde d_t, $t = 2, ..., T$, je délka jednotlivých časových intervalů sledování daného okamžikového ukazatele.

Příklad 1.1

Vypočítejte průměrnou hodnotu roční časové řady hrubého domácího produktu České republiky (HDP_CR) v mld. Kč (srovnatelné ceny), v letech 1994 - 2000:

Tab. 1.1. HDP ČR v letech 1994 - 2000

Rok	HDP ČR
1994	1303,6
1995	1381,1
1996	1447,7
1997	1432,8
1998	1401,3
1999	1390,6
2000	1433,8
Celkem	9790,9

Časová řada hrubého domácího produktu České republiky v letech 1994 - 2000 je intervalovou časovou řadou, proto se její průměrná hodnota vypočítá pomocí aritmetického průměru (1.1)

$$\bar{y} = \frac{9\,790{,}9}{7} = 1\,398{,}7\,.$$

Průměrná hodnota časové řady v letech 1994 až 2000 je 1 398,7 mld. Kč.

Příklad 1.2

Vypočítejte průměrný počet registrovaných uchazečů o zaměstnání v České republice (Nezam_CR) v roce 2000 v tis. osob.

Tab. 1.2: Počet registrovaných uchazečů o zaměstnání v ČR v roce 2000 v tis. osob

Datum	Počet uchazečů	Délka intervalu
31.12.1999	487,6	
31.1.2000	508,5	31
29.2.2000	506,1	29
31.3.2000	493,4	31
30.4.2000	471,2	30
31.5.2000	453,8	31
30.6.2000	451,4	30
31.7.2000	469,7	31
31.8.2000	467,3	31
30.9.2000	458,3	30
31.10.2000	445,2	31
30.11.2000	442,2	30
31.12.2000	457,4	31

Počet registrovaných uchazečů o zaměstnání v České republice v jednotlivých měsících roku 2000 je okamžiková časová řada, protože rozhodným okamžikem sledování je vždy poslední den daného měsíce. Protože vzdálenost mezi jednotlivými okamžiky sledování není stejná, použijeme pro výpočet průměrné hodnoty vážený chronologický průměr (1.3)

$$\bar{y} = \frac{\frac{487{,}6 + 508{,}5}{2} \cdot 31 + \frac{508{,}5 + 506{,}1}{2} \cdot 29 + \ldots + \frac{442{,}2 + 457{,}4}{2} \cdot 31}{31 + 29 + \ldots + 31} = \frac{171\,971{,}6}{366} = 469{,}868\,.$$

Průměrný počet registrovaných uchazečů o zaměstnání v České republice v roce 2000 byl 469,9 tis. osob.

1.3.2 Míry dynamiky

Jednoduché míry dynamiky časových řad umožňují charakterizovat základní rysy "chování" časových řad a formulovat jistá kritéria pro jejich modelování.

Předpokládejme časovou řadu y_t, $t = 1, \ldots, T$. Nejjednodušší mírou dynamiky je *absolutní přírůstek* (první diference), který lze zapsat jako

$$\Delta y_t = y_t - y_{t-1}, \qquad t = 2, \ldots, T. \qquad (1.4)$$

Tato charakteristika vyjadřuje změnu hodnoty v čase t proti času t -1. Často se používá také *průměrný absolutní přírůstek*

$$\bar{\Delta} = \frac{(y_2 - y_1) + (y_3 - y_2) + \ldots + (y_T - y_{T-1})}{T-1} = \frac{\sum_{t=2}^{T} \Delta\, y_t}{T-1} = \frac{y_T - y_1}{T-1}\,. \qquad (1.5)$$

Diferencováním první diference lze získat druhou diferenci, tj. $\Delta^2 y_t = \Delta y_t - \Delta y_{t-1}$, $t = 3, ..., T$, diferencováním druhé diference dostaneme diferenci třetí, tj. $\Delta^3 y_t = \Delta^2 y_t - \Delta^2 y_{t-1}$, $t = 4, ..., T$ atd. Diferencování má v analýze časových řad velký význam. Používá se při modelování trendu časových řad k výběru vhodné trendové funkce, nezastupitelná je jeho role při stochastickém modelování časových řad.

Velmi důležitou mírou dynamiky časových řad je *koeficient růstu*

$$k_t = \frac{y_t}{y_{t-1}}, \qquad t = 2, ..., T. \tag{1.6}$$

Jestliže se tento koeficient vynásobí stem, udává na kolik procent hodnoty v čase t - 1 vzrostla hodnota v čase t. Někdy se pro tento koeficient používá název *tempo růstu*.

Průměrný koeficient růstu (průměrné tempo růstu) se vypočítá jako geometrický průměr*) jednotlivých koeficientů růstu

$$\bar{k} = \sqrt[T-1]{k_2 \cdot k_3 \cdot ... \cdot k_T} = \sqrt[T-1]{\frac{y_2}{y_1} \cdot \frac{y_3}{y_2} \cdot ... \cdot \frac{y_T}{y_{T-1}}} = \sqrt[T-1]{\frac{y_T}{y_1}}. \tag{1.7}$$

Koeficienty růstu se kromě přímého použití pro charakterizování dynamiky časové řady používají jako jedno z kritérií pro nalezení vhodné trendové funkce.

Meziroční koeficient růstu je podíl hodnot časové řady ve stejných obdobích (sezónách) v po sobě jdoucích letech. V případě čtvrtletní časové řady má tvar

$$k_{(4),t} = \frac{y_t}{y_{t-4}}, \qquad t = 5, 6, ..., T, \tag{1.8}$$

lze jej vyjádřit také jako součin (čtvrtletních) koeficientů růstu

$$k_{(4),t} = \frac{y_t}{y_{t-1}} \cdot \frac{y_{t-1}}{y_{t-2}} \cdot \frac{y_{t-2}}{y_{t-3}} \cdot \frac{y_{t-3}}{y_{t-4}}.$$

Časová řada čtvrtých odmocnin meziročních koeficientů růstu je řadou klouzavých geometrických průměrů délky 4 koeficientů růstu původní čtvrtletní časové řady.

Další mírou dynamiky časových řad je *relativní přírůstek*

$$\delta_t = \frac{\Delta y_t}{y_{t-1}} = \frac{y_t - y_{t-1}}{y_{t-1}} = \frac{y_t}{y_{t-1}} - 1, \tag{1.9}$$

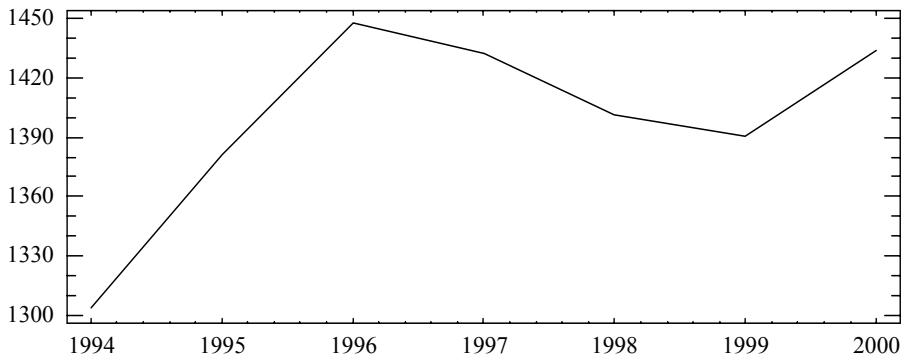
po vynásobení stem nám říká o kolik procent se změnila hodnota časové řady v čase t ve srovnání s časem t - 1. Průměrný relativní přírůstek se vypočítá jako

$$\bar{\delta} = \bar{k} - 1. \tag{1.10}$$

Příklad 1.3

Pro roční časovou řadu reálného hrubého domácího produktu České republiky v letech 1994 - 2000 v mld. Kč (z příkladu 1.1) vypočítejte základní míry dynamiky.

*) Geometrický průměr se používá proto, že smysl má shrnovat koeficienty růstu jejich součinem.



Obr. 1.9: Časová řada hrubého domácího produktu ČR v letech 1994 - 2000 v mld. Kč

V tab. 1.3 jsou na základě vztahů (1.4) až (1.7) vypočítány absolutní přírůstky, koeficienty růstu a relativní přírůstky časové řady.

Tab. 1.3. Míry dynamiky časové řady hrubého domácího produktu ČR

Rok	HDP ČR	Δy_t	k_t	δ_t
1994	1303,6			
1995	1381,1	77,5	1,059	0,059
1996	1447,7	66,6	1,048	0,048
1997	1432,8	-14,9	0,990	-0,010
1998	1401,3	-31,5	0,978	-0,022
1999	1390,6	-10,7	0,992	-0,008
2000	1433,8	43,2	1,031	0,031

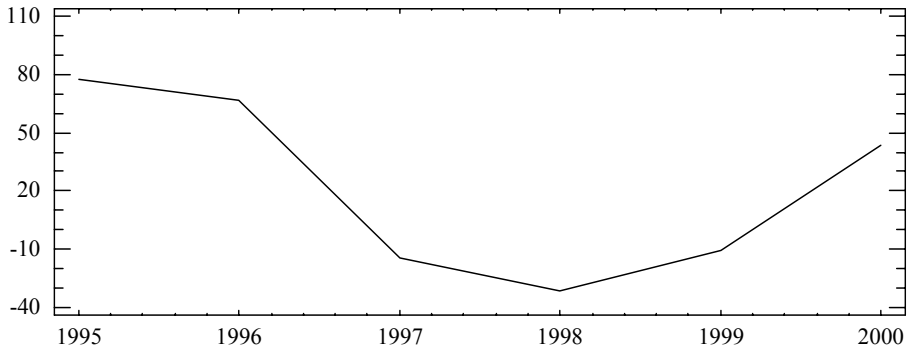
Průměrný absolutní přírůstek

$$\overline{\Delta} = \frac{1\,433{,}8 - 1\,303{,}6}{6} = 21{,}7 \text{ mld. Kč,}$$

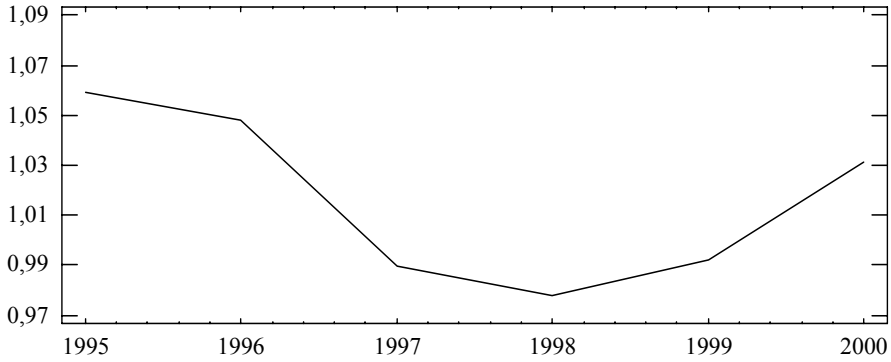
průměrný koeficient růstu

$$\overline{k} = \sqrt[6]{\frac{1\,433{,}8}{1\,303{,}6}} = 1{,}016\,.$$

Na obr. 1.10a je zachycen vývoj absolutních přírůstků, na obr. 1.10b vývoj koeficientů růstu.



Obr. 1.10a: Absolutní přírůstky (I. diference)



Obr. 1.10b: Koeficienty růstu

Výše uvedené výsledky lze získat ve STATGRAPHICSu v panelu **Generate data**.

Postup je následující: v tabulkovém editoru STATGRAPHICSu vybereme prázdný sloupec, klikneme nejprve levým tlačítkem myši na jeho hlavičce (celý se označí) a potom pravým tlačítkem vybereme z nabídky **Generate data**. Výsledky se objeví ve zvoleném sloupci tabulkového editoru po vepsání příslušného příkazu do řádku Expression (obr. 1.11).

Pro výpočet absolutních přírůstků (I. diferencí) lze v řádku Expression použít příkaz vytvořený pomocí operátoru `DIFF`

```
diff(HDP_CR).
```

Koeficienty růstu časové řady vypočítáme pomocí kombinace operátorů `EXP`, `DIFF` a `LOG` příkazem*)

```
exp(diff(log(HDP_CR))).
```

Pro výpočet průměrného absolutního přírůstku použijeme příkaz

```
avg(diff(HDP_CR))
```

a průměrný koeficient růstu získáme příkazem

```
geomean(exp(diff(log(HDP_CR)))).
```

hdp_sc.sf3

	HDP_CR	abs. prirustky	koef. rustu	prum. abs. prir.	prum. koef. rustu
1	1303,6			21,7	1,01599297261
2	1381,1	77,5	1,05945075176		
3	1447,7	66,6	1,0482224314		
4	1432,8	-14,9	0,989707812392		
5	1401,3	-31,5	0,978015075377		
6	1390,6	-10,7	0,992364233212		
7	1433,8	43,2	1,03106572702		
8					

Obr. 1.11: Vypočítané míry dynamiky ve STATGRAPHICSu

*) Zlogaritmujeme-li výraz $k_t = y_t / y_{t-1}$,

dostaneme $\ln k_t = \ln y_t - \ln y_{t-1}$.

Pro grafické zobrazení absolutních přírůstků (obr. 1.10a) a koeficientů růstu (obr. 1.10b) použijeme Horizontal Time Sequence Plot z nabídky v proceduře **Descriptive Methods**, kde do řádku Data vepíšeme příslušný příkaz (stejně jako do řádku Expression v panelu **Generate Data**). Zobrazené grafy je však třeba upravit tak, že nastavíme časovou osu od roku 1995.

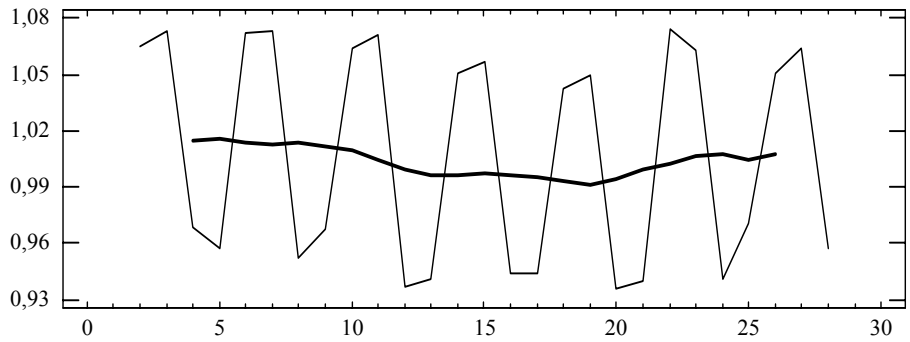
Příklad 1.4

Na základě čtvrtletní časové řady reálného hrubého domácího produktu ČR (HDP_c_CR) od I. čtvrtletí roku 1994 do IV. čtvrtletí roku 2000 v mld. Kč vypočítejte koeficienty růstu (čtvrtletní), meziroční koeficienty růstu a centrované klouzavé geometrické průměry délky 4 počítané z koeficientů růstu.

Tab. 1.4. Koeficienty růstu (čtvrtletní), meziroční koeficienty růstu a centrované klouzavé geometrické průměry reálného HDP ČR

		HDP ČR čtvrtletní	Koeficienty růstu (čtvrtletní) $k_t = y_t/y_{t-1}$	Meziroční koef. růstu $k_{(4),t} = y_t/y_{t-4}$	Centrované klouzavé geometrické průměry $\left(\sqrt[4]{k_{(4),t}} \cdot \sqrt[4]{k_{(4),t-1}}\right)^{1/2}$
1994	I.	302,2			
	II.	321,8	1,065		
	III.	345,2	1,073		
	IV.	334,4	0,969		1,015
1995	I.	319,9	0,957	1,059	1,016
	II.	343,0	1,072	1,066	1,014
	III.	367,9	1,073	1,066	1,013
	IV.	350,3	0,952	1,048	1,014
1996	I.	339,1	0,968	1,060	1,012
	II.	360,7	1,064	1,052	1,010
	III.	386,2	1,071	1,050	1,004
	IV.	361,7	0,937	1,033	0,999
1997	I.	340,4	0,941	1,004	0,996
	II.	357,6	1,051	0,991	0,996
	III.	378,0	1,057	0,979	0,997
	IV.	356,8	0,944	0,986	0,996
1998	I.	336,8	0,944	0,989	0,995
	II.	351,0	1,042	0,982	0,993
	III.	368,6	1,050	0,975	0,991
	IV.	344,9	0,936	0,967	0,994
1999	I.	324,2	0,940	0,963	0,999
	II.	348,1	1,074	0,992	1,002
	III.	370,0	1,063	1,004	1,007
	IV.	348,3	0,941	1,010	1,008
2000	I.	338,2	0,971	1,043	1,005
	II.	355,5	1,051	1,021	1,008
	III.	378,2	1,064	1,022	
	IV.	361,9	0,957	1,039	

Z hlediska symboliky koeficientů růstu je logické, že první hodnota časové řady meziročních koeficientů růstu je umístěna do 1. čtvrtletí druhého roku, tj. roku 1995. Vzhledem k tomu, že po odmocnění čtvrtou odmocninou lze tuto řadu chápat jako řadu klouzavých geometrických průměrů koeficientů růstu, je vhodné umístit její první hodnotu mezi třetí a čtvrté čtvrtletí roku 1994. Centrováním geometrickým průměrem sousedních dvojic hodnot se první hodnota centrovaného klouzavého geometrického průměru dostane do čtvrtého čtvrtletí roku 1994 a poslední hodnota do druhého čtvrtletí roku 2000.



Obr. 1.12: Koeficienty růstu (čtvrtletní) a centrované klouzavé geometrické průměry délky 4

1.3.3 Cvičení

1. Znázorněte charakter čtvrtletní časové řady hrubého domácího produktu České republiky (HDP_c_CR) vhodnými grafy.
2. Vypočítejte průměrnou hodnotu okamžikové časové řady počtu nově registrovaných uchazečů o zaměstnání v České republice v roce 1999 (NezNov_CR).
3. Vypočítejte míry dynamiky časové řady počtu živě narozených dětí v České republice (ZivNar_CR) v letech 1920 - 1999. Znázorněte graficky.
4. Porovnejte pomocí průměrného koeficientu růstu dynamiku časových řad počtu nově registrovaných uchazečů o zaměstnání v České a Slovenské republice (NezNov_CR a NezNov_SR) v roce 1997 a v roce 1999.

2 MODELOVÁNÍ TRENDU A SEZÓNNÍ SLOŽKY

2.1 Dekompozice časové řady

Klasická analýza ekonomických časových řad vychází z předpokladu, že časovou řadu y_t pro $t = 1, 2, \ldots, T$ je možné rozložit na čtyři složky: trendovou, cyklickou, sezónní a nesystematickou.

Trendová složka (T_t) vyjadřuje dlouhodobou tendenci vývoje zkoumaného jevu. Je výsledkem faktorů, které dlouhodobě působí stejným směrem např. technologie výroby, demografické podmínky, podmínky na trhu apod.

Cyklická složka (C_t) vyjadřuje kolísání okolo trendu, ve kterém se střídají fáze růstu a poklesu. Jednotlivé cykly (periody) se vytvářejí za období delší než jeden rok a mají nepravidelný charakter, tj. různou délku a amplitudu. Cykly jsou v ekonomických časových řadách způsobeny ekonomickými a neekonomickými faktory. V posledních letech se věnuje pozornost zejména technologickým, inovačním či demografickým cyklům.

Sezónní složka (S_t) vyjadřuje pravidelné kolísání okolo trendu v rámci kalendářního roku. Sezónní výkyvy se opakují každoročně ve stejných obdobích (délka periody je jeden rok) a vznikají v důsledku střídání ročních období nebo vlivem různých institucionalizovaných zvyků, jako jsou např. svátky, dovolené apod.

Poslední složkou časové řady je *nesystematická složka* (I_t nebo a_t). Tato složka vyjadřuje nahodilé a jiné nesystematické výkyvy, ale také chyby měření apod. Označení I_t se používá při sezónní dekompozici. V regresních modelech a ARIMA modelech budeme nesystematickou složku označovat jako a_t.

Trendová a cyklická složka mohou být přítomné v časových řadách ročních údajů, ale také v krátkodobých časových řadách, tj. v řadách s intervalem sledování kratším než jeden rok, např. čtvrtletních, měsíčních, týdenních, denních apod. Sezónní složka se vyskytuje pouze v krátkodobých časových řadách, obvykle v měsíčních a čtvrtletních. Nesystematická složka je přítomná v každé časové řadě.

Dekompozice časové řady může být:

a) aditivní, hodnoty časové řady se dají určit jako součet hodnot jednotlivých složek, tj.

$$y_t = T_t + C_t + S_t + I_t, \tag{2.1}$$

b) multiplikativní, hodnoty časové řady se dají určit jako součin hodnot jednotlivých složek

$$y_t = T_t \,.\, C_t \,.\, S_t \,.\, I_t. \tag{2.2}$$

Po aditivní dekompozici jsou jednotlivé složky časové řady ve stejných měrných jednotkách jako původní časová řada. Aditivní dekompozice se používá v případě, že variabilita hodnot časové řady je přibližně konstantní v čase.

Po multiplikativní dekompozici je trendová složka časové řady ve stejných měrných jednotkách jako původní časová řada, ale ostatní složky časové řady (cyklická, sezónní a nesystematická) jsou v relativním vyjádření. Multiplikativní dekompozice se používá v případě, že variabilita časové řady roste v čase, nebo se v čase mění. V praxi se dekompozice časových řad často používá z těchto důvodů:

a) analýzou jednotlivých složek řady lze odhalit určité zákonitosti vývoje zkoumaného jevu,
b) časové řady je možné očistit od sezónnosti, tj. z časové řady se odstraní sezónní složka, což umožňuje porovnávat trend několika časových řad současně,
c) časové řady lze očistit od trendu, tj. z řady se odstraní trendová složka, což umožňuje lépe modelovat sezónnost, protože charakter sezónnosti je výraznější,
d) často umožňuje přesněji určit předpovědi nejen jednotlivých složek časové řady, ale v konečném důsledku také samotné časové řady, v tom smyslu, že předpovědi jednotlivých složek se sečtou anebo vynásobí podle toho, který typ dekompozice jsme použili.

2.2 Základní metody modelování trendové složky, konstrukce předpovědí

2.2.1 Analýza trendu

Trend v časových řadách je možné popsat pomocí trendových funkcí a klouzavých průměrů nebo klouzavých mediánů. Modelování trendu pomocí trendových funkcí se používá, pokud vývoj časové řady odpovídá určité funkci času např. lineární, kvadratické, exponenciální, S-křivky apod. Modelování trendu pomocí klouzavých průměrů nebo pomocí klouzavých mediánů se používá, je-li vývoj řady v důsledku silného vlivu nesystematické složky nerovnoměrný, nebo má extrémní hodnoty.

Při modelování trendu pomocí trendových funkcí vycházíme z následujících předpokladů. Časová řada y_t pro $t = 1, 2, \dots, T$ je řadou uspořádaných hodnot v čase t, které získáme měřením určitého ukazatele ve stejně dlouhých časových intervalech t (např. ročních, čtvrtletních, měsíčních apod.).

Předpokládejme, že časovou řadu y_t pro $t = 1, 2, \dots, T$ je možné zapsat ve tvaru

$$y_t = Y_t + a_t, \tag{2.3}$$

kde Y_t představuje teoretický model systematické složky vývoje ekonomického ukazatele Y v čase t a a_t vyjadřuje nesystematickou složku. V analýze časových řad je model Y_t funkcí času t, tj. $Y_t = f(t)$. Pokud se jedná o časovou řadu pouze s trendovou složkou, potom Y_t vyjadřuje model trendu T_t v čase t. Je-li v časové řadě kromě trendové také sezónní složka, nebo cyklická složka, potom je Y_t kompozicí modelů těchto složek. Protože se v této části zabýváme trendovou složkou, $Y_t = T_t$ má model (2.3) má za předpokladu aditivní dekompozice tvar

$$y_t = Y_t + a_t = T_t + a_t, \qquad t = 1, 2, \dots, T$$

kde

1. T_t je systematická složka a představuje deterministický trend, který lze vyjádřit matematickou funkcí časové proměnné t,
2. a_t je nesystematická složka s vlastnostmi procesu bílého šumu, což znamená, že v každém čase t, platí

$$E(a_t) = 0,\ D(a_t) = \sigma_a^2,\ \mathrm{cov}(a_t, a_{t-k}) = 0 \text{ a } a_t \sim N(0, \sigma_a^2), \tag{2.4}$$

tj. náhodné veličiny a_t mají v čase t nulovou střední hodnotu, konstantní rozptyl, jsou vzájemně lineárně nezávislé a mají normální rozdělení.

Trendové funkce

Konstantní trend má formu

$$T_t = \beta_0, \qquad t = 1, 2, \dots, T, \tag{2.5}$$

tj. hodnoty trendu se vzhledem k časové proměnné t nemění, jsou konstantní.

Odhad parametru β_0 se získá metodou nejmenších čtverců, jako průměr hodnot časové řady, tj.

$$\hat{\beta}_0 = \bar{y} = \frac{1}{T}\sum_{t=1}^{T} y_t \,. \tag{2.6}$$

Odhad trendu v čase t je

$$\hat{T}_t = \hat{y}_t = \bar{y}\,, \tag{2.7}$$

odhad rozptylu nesystematické složky

$$\hat{\sigma}_a{}^2 = \frac{1}{T-1}\sum_{t=1}^{T}(y_t - \bar{y})^2 = \hat{\sigma}_y{}^2 \,. \tag{2.8}$$

Lineární trendová funkce (přímka) má tvar

$$T_t = \beta_0 + \beta_1 t, \qquad t = 1, 2, \dots, T. \tag{2.9}$$

Parametry lineárního trendu β_0, β_1 odhadujeme metodou nejmenších čtverců podle vztahů

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{t} \qquad \hat{\beta}_1 = \frac{\sum_{t=1}^{T} t y_t - \bar{t}\sum_{t=1}^{T} y_t}{\sum_{t=1}^{T} t^2 - T(\bar{t})^2}, \tag{2.10}$$

kde

$$\bar{t} = \frac{1}{T}\sum_{t=1}^{T} t = \frac{T+1}{2}.$$

Odhad lineárního trendu je

$$\hat{T}_t = \hat{y}_t = \hat{\beta}_0 + \hat{\beta}_1 t\,, \tag{2.11}$$

odhad rozptylu nesystematické složky lineárního trendu

$$\hat{\sigma}_a{}^2 = \frac{1}{T-2}\sum_{t=1}^{T}(y_t - \hat{y}_t)^2 \,. \tag{2.12}$$

Kvadratická trendová funkce (parabola) má tvar

$$T_t = \beta_0 + \beta_1 t + \beta_2 t^2 \qquad t = 1, 2, \dots, T. \tag{2.13}$$

Parametry kvadratického trendu β_0, β_1 a β_2 odhadujeme metodou nejmenších čtverců. Odhad kvadratického trendu je

$$\hat{T}_t = \hat{y}_t = \hat{\beta}_0 + \hat{\beta}_1 t + \hat{\beta}_2 t^2\,, \tag{2.14}$$

a odhad rozptylu nesystematické složky kvadratického trendu

$$\hat{\sigma}_a{}^2 = \frac{1}{T-3}\sum_{t=1}^{T}(y_t - \hat{y}_t)^2 \,. \tag{2.15}$$

Exponenciální trendová funkce má tvar

$$T_t = \beta_0 \beta_1^{\,t}, \qquad t = 1, 2, \dots, T. \qquad (2.16)$$

Parametry exponenciálního trendu β_0, $\beta_1 > 0$ odhadujeme metodou nejmenších čtverců, po linearizující transformaci logaritmováním, tj.

$$\ln T_t = \ln \beta_0 + t \ln \beta_1, \qquad t = 1, 2, \dots, T. \qquad (2.17)$$

Odhad exponenciálního trendu je

$$\hat{T}_t = \hat{y}_t = \hat{\beta}_0 \hat{\beta}_1^{\,t}, \qquad (2.18)$$

a odhad rozptylu nesystematické složky exponenciálního trendu (po linearizaci logaritmováním)

$$\hat{\sigma}_a^{\,2} = \frac{1}{T-2} \sum_{t=1}^{T} (y_t - \hat{y}_t)^2 . \qquad (2.19)$$

S-křivka má tvar

$$T_t = e^{(\beta_0 + \beta_1 \frac{1}{t})}, \qquad t = 1, 2, \dots, T. \qquad (2.20)$$

S-křivka po logaritmické transformaci má tvar hyperboly

$$\ln T_t = \beta_0 + \beta_1 \frac{1}{t}, \qquad t = 1, 2, \dots, T. \qquad (2.21)$$

Parametry modelu (2.21) odhadujeme metodou nejmenších čtverců.

Odhad *S*-křivky je

$$\hat{T}_t = \hat{y}_t = e^{(\hat{\beta}_0 + \hat{\beta}_1 \frac{1}{t})},$$

odhad rozptylu nesystematické složky *S*-křivky (po linearizaci logaritmováním)

$$\hat{\sigma}_a^{\,2} = \frac{1}{T-2} \sum_{t=1}^{T} (y_t - \hat{y}_t)^2 . \qquad (2.22)$$

Modifikovaný exponenciální trend má tvar

$$T_t = \gamma + \beta_0 \beta_1^{\,t}, \qquad t = 1, 2, \dots, T, \qquad (2.23)$$

pro $\beta_0 < 0$, $0 < \beta_1 < 1$ a $\gamma > 0$. Konstanta γ je asymptotou (úrovní saturace, hladinou nasycení), ke které trend časové řady pro $t \to \infty$ konverguje. Přírůstek exponenciálního trendu β_1 je pomalejší, než přírůstek lineárního trendu.

Modifikovaný exponenciální trend je populární v marketingu, ve kterém se využívají trendové funkce s horním omezením. Je to nelineární funkce, kterou není možné linearizovat žádnou transformací a proto se její parametry odhadují iterativními metodami např. Marquartovou nebo Gaussovou-Newtonovou iterativní metodou. Tyto metody vyžadují výpočet počátečních odhadů parametrů funkce, které se dají získat např. metodou částečných součtů, metodou vybraných bodů apod.

Logistický trend uvádíme ve tvaru Pearlovy-Reedovy trendové funkce

$$T_t = \frac{1}{\gamma + \beta_0 \beta_1^{\,t}} \qquad t = 1, 2, \dots, T. \qquad (2.24)$$

jejíž inverzní funkce $1/T_t = \gamma + \beta_0 \beta_1^t$ má tvar modifikovaného exponenciálního trendu. Parametry modelu (2.24) se po inverzní transformaci odhadují stejným způsobem, jako pro modifikovaný exponenciální trend.

Gompertzův trend má tvar

$$T_t = \gamma \beta_0^{\beta_1^t}, \qquad t = 1, 2, \dots, T. \qquad (2.25)$$

a po logaritmické transformaci má tvar modifikovaného exponenciálního trendu

$$\ln T_t = \ln \gamma + \beta_1^t \ln \beta_0 \quad \text{resp.} \quad T_t^* = \gamma^* + \beta_0^* \beta_1^t. \qquad (2.26)$$

Křivka má horní asymptotu $\gamma^* = \ln \gamma$ a vyjadřuje hranici nasycení pro $t \to \infty$, např. prodeje nebo spotřeby určitého výrobku na trhu.

Parametry modelu (2.25) se po transformaci (2.26) odhadují jako u modifikovaného exponenciálního trendu nebo jednoduchého exponenciálního trendu. Po úpravě (2.26) dostaneme $T_t^* - \gamma^* = \alpha^* \beta^t$. Jestliže označíme $(T_t^* - \gamma^*) = T_t^{**}$, potom po další logaritmické transformaci lze odhadnout α^* a β metodou nejmenších čtverců, jako parametry lineárního trendu.

Bodové a intervalové předpovědi – extrapolace

V období analýzy nebo interpolace řady $t = 1, 2, \dots, T$ zjišťujeme, zda řada y_t má trend a hledáme jeho model. Z hodnot časové řady y_t potom odhadneme parametry modelu. Je-li odhad trendu statisticky významný, využijeme jej jako prognostický model pro výpočet předpovědí - extrapolací.

Extrapolacemi nazýváme kvantitativní odhady budoucích hodnot časové řady, které vznikají prodloužením vývoje z minulosti a přítomnosti do budoucnosti s horizontem h za předpokladu, že se tento vývoj nezmění. Extrapolační předpovědi rozdělujeme na bodové a intervalové.

Bodová předpověď - extrapolace "ex ante" se určuje v čase $t = T$ (začátek předpovídání, práh predikce) do horizontu h, tj. do časového bodu $t = T + h$ a označujeme ji $\hat{y}_T(h)$. Horizontem předpovídání se rozumí počet období $h > 0$ od bodu $t = T$ do budoucnosti.

(1 - α) x 100% interval předpovědí je interval, ve kterém se s pravděpodobností (1 - α) x 100% nachází skutečná hodnota y_{T+h}, tj.

$$\hat{y}_T(h) \pm t_{1-\alpha/2}(T - l - 1)\, \hat{\sigma}_p ,$$

kde $t_{1-\alpha/2}(T - l - 1)$ je (1 - α/2) x 100% kvantil Studentova rozdělení s $T - (l + 1)$ stupni volnosti, kde $l + 1$ je počet odhadnutých parametrů v polynomiálních funkcích, $\hat{\sigma}_p$ je směrodatná chyba předpovědi v horizontu h.

Když určujeme extrapolace předpokládáme, že

1. vybraný model je správný,
2. skutečné parametry modelu se v čase nemění.

V mnoha situacích jsou tyto předpoklady nereálné, protože proces, který generuje vývoj časové řady se mění v čase. Očekávaná přesnost předpovědí proto závisí na horizontu předpovědi h. Čím je horizont předpovědí delší, tím je možné očekávat větší chyby předpovědí.

Chyba předpovědi a její rozptyl

Při extrapolaci se dopouštíme chyby, která je dána vztahem

$$\hat{a}_{T+h} = y_{T+h} - \hat{y}_T(h), \qquad h > 0. \tag{2.27}$$

Chybu předpovědi (2.27) lze rozložit na dvě složky

$$\hat{a}_{T+h} = (y_{T+h} - Y_{T+h}) + (Y_{T+h} - \hat{y}_T(h)), \tag{2.28}$$

kde (y_{T+h} - Y_{T+h}) je chyba způsobená volbou modelu (předpokládáme, že model je zvolen správně, tj. y_{T+h} - Y_{T+h} = 0) a $(Y_{T+h} - \hat{y}_T(h))$ je chyba způsobená odhadem parametrů modelu.

Požadujeme, aby bodová předpověď byla

1. nezkresleným odhadem

$$E\{(\hat{a}_{T+h})\} = E\{(y_{T+h} - \hat{y}_T(h))\} = 0\,, \tag{2.29}$$

2. vydatným odhadem

$$E\{(\hat{a}_{T+h})^2\} = E\{(y_{T+h} - \hat{y}_T(h))^2\} = \sigma_p^2 \Rightarrow \min. \tag{2.30}$$

Rozptyl chyby předpovědi (2.30) je možné odvodit pro různé trendové funkce (viz např. Montgomery, D. C. (1990), str. 172 - 184).

Na závěr uvedeme bodové předpovědi a předpovědní intervaly pro vybrané trendové funkce.

Mějme časovou řadu y_t, $t = 1, 2, \ldots, T$ na jejímž základě určíme předpovědi pro horizont $h > 0$, za předpokladu následujících trendů.

Konstantní trend

$$y_t = \beta_0 + a_t,$$

bodová předpověď

$$\hat{y}_T(h) = \hat{\beta}_0 = \bar{y}\,, \tag{2.31}$$

(1 - α) x 100% předpovědní interval

$$\bar{y} \pm \hat{\sigma}_a t_{1-\alpha/2}(T-1)\sqrt{1+\frac{1}{T}}\,, \tag{2.32}$$

kde $t_{1-\alpha/2}$ je (1 - α/2) x 100% kvantil Studentova rozdělení s (T - 1) stupni volnosti, $\hat{\sigma}_a$ je odhad směrodatné odchylky nesystematické složky (směrodatná odchylka reziduí).

Lineární trend

$$y_t = \beta_0 + \beta_1 t + a_t,$$

bodová předpověď

$$\hat{y}_T(h) = \hat{\beta}_0 + \beta_1(T+h)\,, \tag{2.33}$$

(1 - α) x 100% předpovědní interval

$$\hat{y}_T(h) \pm \hat{\sigma}_a t_{1-\alpha/2}(T-2)\sqrt{1+\frac{1}{T}+\frac{(T+h-\bar{t})^2}{T(T^2-1)/12}}\,, \tag{2.34}$$

kde $t_{1-\alpha/2}$ je (1 - α/2) x 100% kvantil Studentova rozdělení s (T - 2) stupni volnosti, $\hat{\sigma}_a$ je odhad směrodatné odchylky nesystematické složky (směrodatná odchylka reziduí) lineárního trendu.

Kvadratický trend

$$y_t = \beta_0 + \beta_1 t + \beta_2 t^2 + a_t,$$

bodová předpověď

$$\hat{y}_T(h) = \hat{\beta}_0 + \hat{\beta}_1(T+h) + \hat{\beta}_2(T+h)^2, \tag{2.35}$$

(1 - α) x 100% předpovědní interval

$$\hat{y}_T(h) \pm \hat{\sigma}_a t_{1-\alpha/2}(T-3)\sqrt{1+\{1,(T+h),(T+h)^2\}(\boldsymbol{X}'\boldsymbol{X})^{-1}\{1,(T+h),(T+h)^2\}'}, \tag{2.36}$$

kde $t_{1-\alpha/2}$ je (1 - α/2) x 100% kvantil Studentova rozdělení s (T - 3) stupni volnosti, $\hat{\sigma}_a$ je odhad směrodatné odchylky nesystematické složky (sm. odchylka reziduí) paraboly a

$$\mathbf{X} = \begin{bmatrix} 1, & 1, & 1 \\ 1, & 2, & 4, \\ \vdots & \vdots & \vdots \\ 1, & T, & T^2 \end{bmatrix}.$$

Ověřování vhodnosti trendové funkce

Výběr trendové funkce (nebo jiného modelu trendu časové řady) provádíme na základě:

- grafu časové řady nebo jejích absolutních či relativních charakteristik,
- interpolačních kritérií (směrodatná odchylka reziduí, koeficient determinace, koeficient autokorelace reziduí, testy parametrů),
- extrapolačních kritérií (průměrné charakteristiky chyb předpovědí "ex post", graf předpověď – skutečnost).

Grafická analýza

Na začátku analýzy provádíme předběžný výběr trendové funkce pomocí grafu časové řady nebo pomocí grafické analýzy diferencí a koeficientů růstu daných časových řad. Je známo, že pokud

1. řada prvních diferencí (y_t - y_{t-1}) pro t = 2, 3, ..., T kolísá okolo nuly, volíme konstantní trend,
2. řada prvních diferencí (y_t - y_{t-1}) pro t = 2, 3, ..., T kolísá okolo nenulové konstanty, volíme lineární trend,
3. řada prvních diferencí (y_t - y_{t-1}) pro t = 2, 3, ..., T má přibližně lineární trend a řada druhých diferencí (y_t - 2y_{t-1} + y_{t-2}) pro t = 3, 4, ..., T má přibližně konstantní trend, volíme kvadratický trend (parabolu),

4. řada koeficientů růstu y_t/y_{t-1} pro $t = 2, 3, ..., T$ nebo řada prvních diferencí ($\ln y_t - \ln y_{t-1}$) kolísá okolo nenulové konstanty, volíme jednoduchý exponenciální trend,
5. řada $\ln y_t$ pro $t = 1, 2, ..., T$ má přibližně hyperbolický průběh, volíme *S*-křivku,
6. řada podílů sousedních diferencí $(y_t - y_{t-1})/(y_{t-1} - y_{t-2})$ pro $t = 3, 4, ..., T$ kolísá okolo nenulové konstanty, volíme modifikovaný exponenciální trend,
7. řada podílů sousedních diferencí $(\ln y_t - \ln y_{t-1})/(\ln y_{t-1} - \ln y_{t-2})$ pro $t = 3, 4, ..., T$ kolísá okolo nenulové konstanty, volíme Gompertzovu křivku.

Výběr trendové funkce na základě grafu je subjektivní a v případě složitějších funkcí nebo mají-li časové řady velkou variabilitu, nevede k jednoznačným výsledkům.

Druhý způsob výběru trendové funkce je objektivnější, protože se zakládá na matematicko-statistických kritériích dvojího charakteru:

1. interpolační kritéria (průměrné charakteristiky reziduí, Durbinova–Watsonova statistika, reziduální autokorelační funkce, statistická významnost parametrů trendu, index determinace a upravený index determinace),
2. extrapolační kritéria (míry přesnosti předpovědí "ex post" a Theilův koeficient nesouladu).

Interpolační kritéria

Po odhadu parametrů modelu trendu z časové řady y_t, pro $t = 1, 2, ... , T$, např. metodou nejmenších čtverců zjišťujeme, jak přesně tento model vystihuje skutečnou časovou řadu, tj. zkoumáme charakter rozdílů skutečných hodnot y_t určitého ukazatele a vyrovnaných $\hat{y}_t$, resp. odhadnutých hodnot trendu $\hat{T}_t$, tohoto ukazatele v čase $t = 1, 2, ... , T$. Rozdílům $y_t - \hat{y}_t = y_t - \hat{T}_t = \hat{a}_t$ říkáme *rezidua* a jsou odhadem nesystematické složky a_t čase $t = 1, 2, ... , T$.

Přesnost vyrovnávání časové řady y_t, pro $t = 1, 2, ... , T$ měříme průměrnými reziduálními charakteristikami, které lze zobecnit pro libovolný model časové řady (nejen pro trendové funkce).

Míry přesnosti vyrovnávání nebo průměrné charakteristiky reziduí

Průměrná chyba

$$ME = \frac{1}{T}\sum_{t=1}^{T}(y_t - \hat{y}_t) = \frac{1}{T}\sum_{t=1}^{T}\hat{a}_t \ , \tag{2.37}$$

Průměrná čtvercová chyba - rozptyl

$$MSE = \frac{1}{T}\sum_{t=1}^{T}(y_t - \hat{y}_t)^2 = \frac{1}{T}\sum_{t=1}^{T}\hat{a}_t^{\,2} \ , \tag{2.38}$$

Průměrná absolutní chyba

$$MAE = \frac{1}{T}\sum_{t=1}^{T}\left|y_t - \hat{y}_t\right| = \frac{1}{T}\sum_{t=1}^{T}\left|\hat{a}_t\right|, \tag{2.39}$$

Průměrná absolutní procentuální chyba

$$MAPE = \frac{1}{T}\sum_{t=1}^{T}\frac{|y_t - \hat{y}_t|}{y_t}.100 = \frac{1}{T}\sum_{t=1}^{T}\frac{|\hat{a}_t|}{y_t}.100\,, \tag{2.40}$$

Průměrná procentuální chyba

$$MPE = \frac{1}{T}\sum_{t=1}^{T}\frac{(y_t - \hat{y}_t)}{y_t}.100 = \frac{1}{T}\sum_{t=1}^{T}\frac{\hat{a}_t}{y_t}.100\,. \tag{2.41}$$

Zvolená trendová funkce je tím lepší, čím nižší jsou hodnoty uvedených charakteristik.

Durbinův – Watsonův test

Nekorelovanost v nesystematické složce (nepřítomnost autokorelace nesystematické složky) testujeme pomocí prvního koeficientu autokorelace

H_0: $\rho_1 = 0$ autokorelace není, tj. $\mathrm{cov}(a_t, a_{t-1}) = 0$,

H_1: $\rho_1 \neq 0$ autokorelace.

Durbinovo-Watsonovo kritérium má tvar

$$DW = \frac{\sum_{t=2}^{T}(\hat{a}_t - \hat{a}_{t-1})^2}{\sum_{t=1}^{T}\hat{a}_t^2} \tag{2.42}$$

a může nabýt hodnot z intervalu ⟨0, 4⟩.

Rozhodnutí o nezamítnutí nebo zamítnutí testované hypotézy na 5% hladině významnosti vyžaduje určení kritických hodnot, které jsou uvedeny v příloze. Výsledky Durbinova –Watsonova testu jsou v tab. 2.1.

Tab. 2.1. Výsledky Durbinova-Watsonova testu

DW	Výsledek
$4 - d_L < \mathrm{DW} < 4$	H_0 se zamítá - autokorelace
$4 - d_U < \mathrm{DW} < 4 - d_L$	Neumíme rozhodnout, je třeba zvýšit T
$2 < \mathrm{DW} < 4 - d_U$	Přijímá se H_0 - autokorelace není
$d_U < \mathrm{DW} < 2$	Přijímá se H_0 - autokorelace není
$d_L < \mathrm{DW} < d_U$	Neumíme rozhodnout, je třeba zvýšit T
$0 < \mathrm{DW} < d_L$	H_0 se zamítá - autokorelace

Reziduální autokorelační funkce

Mírou lineární závislosti časově zpožděných veličin a_t a a_{t-k} jsou koeficienty autokorelace reziduí definované vztahem

$$r_k = \hat{\rho}_k = \frac{\sum_{t=k+1}^{T} \hat{a}_t \, \hat{a}_{t-k}}{\sum_{t=1}^{T} \hat{a}_t^2} \in \langle -1, 1 \rangle. \tag{2.43}$$

Graf, ve kterém jsou na vodorovné ose časová zpoždění a na svislé ose koeficienty autokorelace reziduí r_k se nazývá *reziduální autokorelační funkcí (ACF)*. Pokud žádný autokorelační koeficient r_k nepřekračuje meze 95% intervalu spolehlivosti $(-2/\sqrt{T}, 2/\sqrt{T})$, je možné předpokládat, že nesystematická složka není autokorelovaná.

Index determinace R^2 a modifikovaný index determinace R^2_M

Index determinace definujeme vztahem

$$R^2 = 1 - \frac{\sum_{t=1}^{T} (y_t - \hat{y}_t)^2}{\sum_{t=1}^{T} (y_t - \bar{y})^2} \in \langle 0, 1 \rangle. \tag{2.44}$$

Čím je hodnota indexu determinace bližší k jedničce (nebo 100 %), tím lépe model vystihuje trend časové řady a naopak.

Nedostatkem koeficientu determinace je, že závisí na počtu parametrů modelu (trendové funkce). Tento nedostatek odstraňuje modifikovaný index determinace ve tvaru

$$R_M^2 = R^2 - \frac{(1 - R^2)(k-1)}{T-k}, \tag{2.45}$$

kde k je počet parametrů modelu (trendové funkce).

t-test pro parametry modelu (trendové funkce)

Pomocí t-testu testujeme hypotézy

H_0: $\beta_i = 0$,

H_1: $\beta_i \neq 0$, pro $i = 0, 1, ..., k$.

Testové kritérium má tvar

$$t = \frac{\hat{\beta}_i}{s_{\hat{\beta}_i}} \sim t(T-k), \tag{2.46}$$

kde $\hat{\beta}_i$ je odhad parametru modelu (trendové funkce), $s_{\hat{\beta}_i}$ je odhad směrodatné chyby odhadu testovaného parametru. Testové kritérium t je náhodná veličina, která má Studentovo t rozdělení s $(T - k)$ stupni volnosti.

Extrapolační kritéria

Extrapolační kritéria se zakládají na principu, že časovou řadu y_t, pro $t = 1, 2, \ldots, T$ rozdělíme na dvě časti. První část řady (testovací část) má T_1 pozorovaní a slouží k výběru modelu trendu, odhadu jeho parametrů a ověření vhodnosti pomocí interpolačních kritérií. Druhá část řady, má délku $(T - T_1)$ pozorování pro $t = T_1 + 1, T_1 + 2, \ldots, T_1 + T_2 = T$ a používá se pro určení předpovědí známé skutečnosti (prognózy "ex post") a pro ověření jejich přesnosti. Přesnost předpovědí "ex post" zhodnotíme pomocí průměrné chyby *ME*, průměrné čtvercové chyby *MSE*, průměrné absolutní procentuální chyby *MAPE* a průměrné procentuální chyby *MPE* podle vztahů (2.37) – (2.41) po příslušné úpravě. Tak např. průměrná chyba předpovědí "ex post"

$$ME = \frac{1}{T_2} \sum_{t=T_1+1}^{T=T_1+T_2} \hat{a}_t(T_1)\,, \tag{2.47}$$

je mírou zkreslení (vychýlení). Pokud je $ME > 0$, model systematicky podhodnocuje skutečnost, pokud je $ME < 0$, model skutečnost systematicky nadhodnocuje.

Jsou-li předpovědi systematicky zkreslené, je vhodné zkoumat statistickou významnost tohoto zkreslení, pomocí testu

$$t(ME) = \frac{ME}{\hat{\sigma}_a / \sqrt{T_2}} \sim t(T_2 - l - 1), \tag{2.48}$$

kde *ME* určíme podle (2.47), l stupeň polynomu trendové funkce a $\hat{\sigma}_a$ je směrodatná odchylka reziduí v období interpolace.

Pokud $|t(ME)| > t_{1-\alpha/2}(T_2 - l - 1)$, zamítáme testovanou hypotézu o nezkreslenosti předpovědí "ex post " na α% hladině významnosti.

Heteroskedasticitu chyb předpovědí (chyby předpovědí rostou v čase) lze ověřit pomocí testového kritéria

$$F = \frac{1}{\hat{\sigma}_a^{\,2} T_2} \sum_{t=T_1+1}^{T=T_1+T_2} \hat{a}_t^2 \sim F(T_2, T_1 - l - 1), \tag{2.49}$$

která má F rozdělení s T_2 a $(T_1 - l - 1)$ stupni volnosti.

Je-li $F > F_{1-\alpha}(T_2, T_1 - l - 1)$, zamítáme testovanou hypotézu a tvrdíme, že chyby předpovědí jsou systematicky zkreslené a model trendu není vhodný pro předpovídání.

Kromě výše uvedených měr, je možné kvalitu předpovědí hodnotit pomocí grafu předpověď – skutečnost, kde se na vodorovnou osu zobrazují hodnoty předpovědí $\hat{y}_t(T_1)$ a na svislou osu skutečné hodnoty y_t. Leží-li skutečné hodnoty a předpovědi na přímce v úhlu 45°, jsou předpovědi bezchybné. V opačném případě, větší vzdálenost bodů od přímky charakterizuje větší zkreslení předpovědí. Jsou-li skutečné hodnoty vyšší než předpovídané (body leží nad přímkou bezchybných předpovědí), potom model podhodnocuje skutečnost. Jsou-li skutečné hodnoty nižší než předpovídané (body leží pod přímkou bezchybných předpovědí), potom model skutečnost nadhodnocuje.

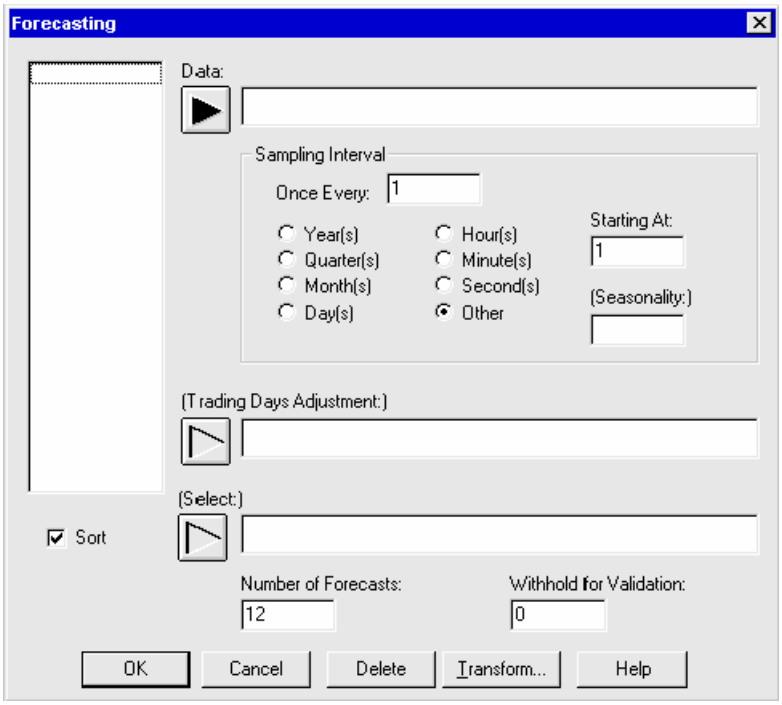
Předpovědi "ex ante"

Jestliže jsme na základě extrapolačních kritérií vybrali vhodný prognostický model, potom tento model aplikujeme na celou časovou řadu y_t, pro $t = 1, 2, \ldots, T$ a určíme předpovědi "ex ante".

2.2.2 Analýza trendu ve STATGRAPHICSu

Analýzu trendu pomocí trendových funkcí provádíme ve STATGRAPHICSu v proceduře **Forecasting**

Special
⇒ **Time-Series Analysis**
⇒ **Forecasting**.

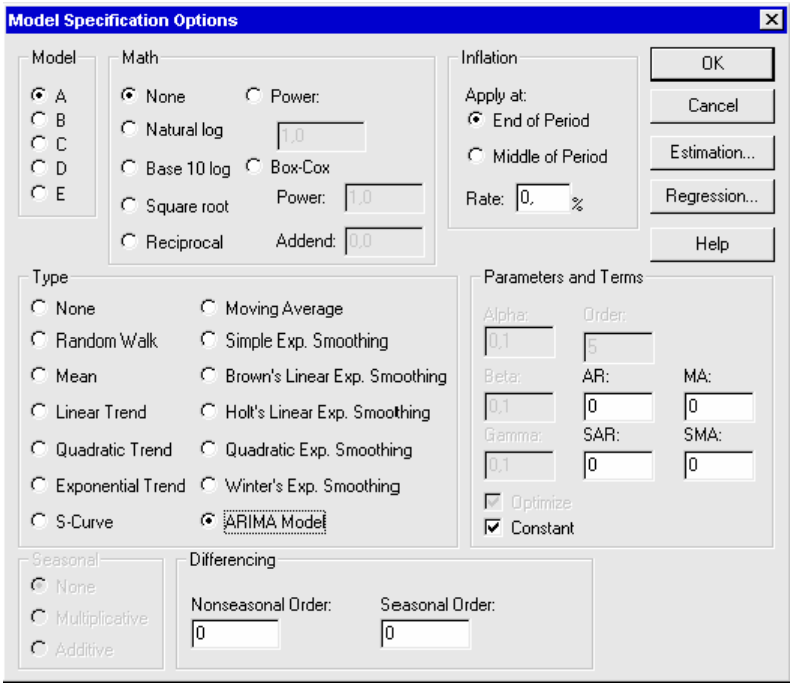


Obr. 2.1: Vstupní panel Forecasting

Ve vstupním panelu vložíme do řádku **Data** název časové řady. V části **Sampling interval** specifikujeme časový interval, za který se zjišťují údaje časové řady. Once Every: 1 znamená, že údaje jsou seřazeny chronologicky za sebou. Do políčka Starting At: zadáme počátek časové řady, do Seasonality: potom počet sezón sezónních časových řad (pro čtvrtletní řady 4, pro měsíční 12 a pod.). Do políčka **Number of forecasts:** zadáme počet předpovědí a do políčka **Withold for Validation:** počet hodnot, které se odtrhnou z konce časové řady a využijí se pro analýzu přesnosti předpovědí "ex post".

Po specifikaci časové řady a jejích vlastností ve vstupním panelu se objeví výsledkové okno Analysis Summary pro zvolený prognostický model. Jako první je pevně

nastavený model Random Walk (stochastický trend). Doplňkový panel ***Model Specification Options*** (obr. 2.2) pro výběr jiného prognostického modelu získáme kliknutím pravého tlačítka myši na okně s výsledky a výběrem *Analysis Options* z doplňkové nabídky.



Obr. 2.2: Doplňkový panel Model Specification Options

V části **Type** lze vybírat z následujících trendů

None	- žádný,
Random Walk	- stochastický trend,
Mean	- konstantní trend,
Linear Trend	- lineární trend,
Quadratic Trend	- kvadratický trend,
Exponential Trend	- exponenciální trend,
S-Curve	- *S*-křivka.

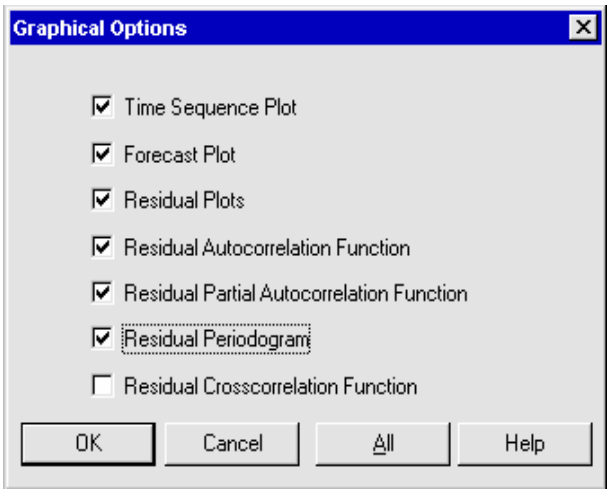
Pokud potřebujeme časovou řadu transformovat můžeme v části **Math** vybrat některou z následujících transformací:

None	- žádná,
Natural log	- logaritmická transformace se základem „e“,
Base 10 log	- logaritmická transformace se základem „10“,
Square root	- transformace druhou odmocninou,
Reciprocal	- transformace převrácenými hodnotami,
Power	- transformace ve formě *r*-tých mocnin
Box-Cox	- Boxova-Coxova transformace.

Pro porovnávání vhodnosti různých modelů trendu, kterými popisujeme trend stejné časové řady, využíváme část **Model.** V této části označíme písmeny jednotlivé

modely v takovém pořadí, v jakém je používáme. Např. lineární trend označíme jako A, kvadratický trend jako B a exponenciální trend jako C.

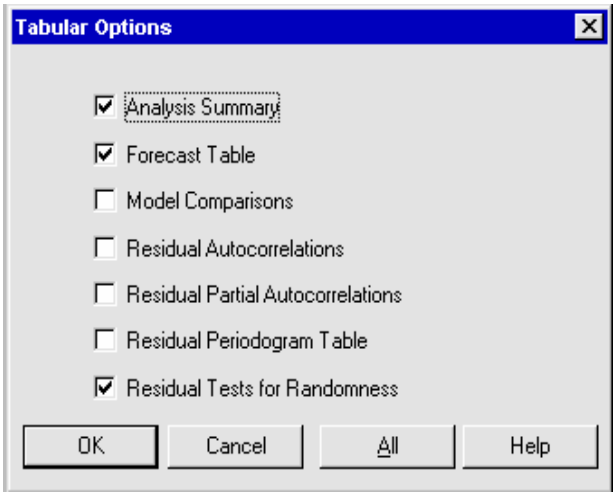
Volbou *Graphical options* získáme výběr grafů, které jsou cennými nástroji analýzy. Z nabídky grafů (obr. 2.3a) využijeme v trendové analýze graf časové řady s vyrovnanými hodnotami a předpověďmi (Time Sequence Plot), graf předpovědí (Forecast Plot), graf reziduí (Residual Plot), graf reziduální autokorelační funkce pro testování autokorelace nesystematické složky (Residual Autocorrelation Function) a periodogram reziduí (Residual Periodogram) využíváme pro zkoumání délky periody v reziduích.



Graf časové řady s předpověďmi
Graf předpovědí
Graf reziduí
Graf reziduální autokorelační funkce
Graf rezid. parciální autokorelační funkce
Graf reziduálního periodogramu
Graf rezid. vzájemné korelační funkce

Obr. 2.3a: Nabídka grafů

Volbou *Tabular options* získáme číselné výsledky analýzy, tj. odhady parametrů modelů, výsledky testů, předpovědi, průměrné charakteristiky reziduí nebo chyb předpovědí "ex post", porovnání výsledků odhadů parametrů různých modelů apod. Nabídka výpočetních metod je na obr. 2.3b.



Odhad modelu a výstupní tabulka
Předpovědi
Porovnání modelů
Reziduální autokorelační funkce
Rezid. parciální autokorelační funkce
Reziduální periodogram
Testy reziduí

Obr. 2.3b: Nabídka výpočetních metod

Z této nabídky pro analýzu trendu využíváme kromě Analysis Summary, tabulku vyrovnaných hodnot a předpovědí (Forecast Table), výsledky porovnávání modelů (Model

Comparisons), hodnoty reziduální autokorelační funkce (Residual Autocorrelations) a testy reziduí (Residual Tests for Randomness).

Pro kvalitní trendovou analýzu je však obvykle třeba použít i jiná než uvedená interpolační kritéria (např. Durbinovu-Watsonovu statistiku a index determinace) nebo jiné modely trendových funkcí, která STATGRAPHICS v této proceduře nenabízí. K tomuto účelu potom využíváme proceduru **Multiple Regression**

Relate
$\Rightarrow$ **Multiple Regression**.

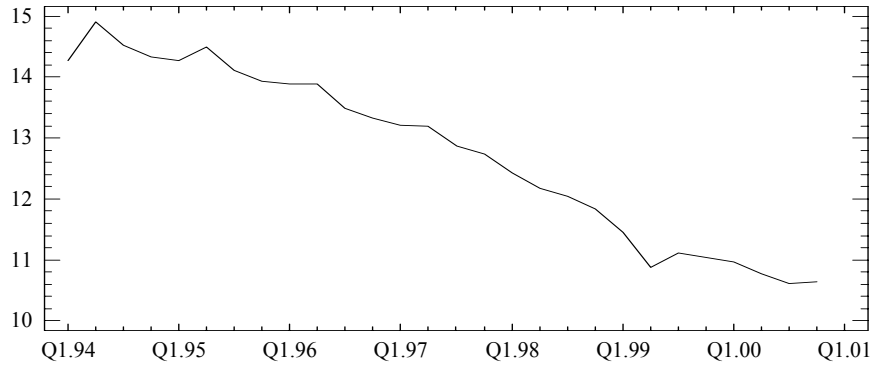
Kde ve vstupním panelu vkládáme do řádku Dependent Variable časovou řadu a do políčka Independent Variables časovou proměnnou. Odhadnutý model trendu a vypočítané hodnoty Durbinovy-Watsonovy statistiky, indexu determinace a modifikovaného indexu determinace získáme ve výsledkovém okně.

Příklady vyplnění vstupního panelu této procedury pro odhady parametrů trendových funkcí nyní uvádíme (y_t značí časovou řadu, t časovou proměnnou).

Lineární trend (přímka)	Dependent Variable:	y_t
	Independent Variables:	t
Kvadratický trend (parabola)	Dependent Variable:	y_t
	Independent Variables:	t
		t^2
Exponenciální trend (exponenciála)	Dependent Variable:	log(y_t)
	Independent Variables:	t
S-křivka	Dependent Variable:	log(y_t)
	Independent Variables:	1/t

Příklad 2.1

Jsou dány sezónně očištěné čtvrtletní údaje o podílu nezaměstnaných 30 – 34 letých na celkové nezaměstnanosti Slovenska v letech 1994 až 2000 v % (SA_PodNez_SR). Cílem analýzy časové řady je vybrat model trendu, ověřit jeho vhodnost (analyzovat rezidua) a určit předpovědi trendu od počátku předpovídání ve čtvrtém čtvrtletí roku 2000 s horizontem h = 1, 2, 3, 4 (tj. na rok 2001).



Obr. 2.4: Sezónně očištěná čtvrtletní časová řada podílu nezaměstnaných 30 – 34 letých na celkové nezaměstnanosti Slovenska v letech 1994 - 2000

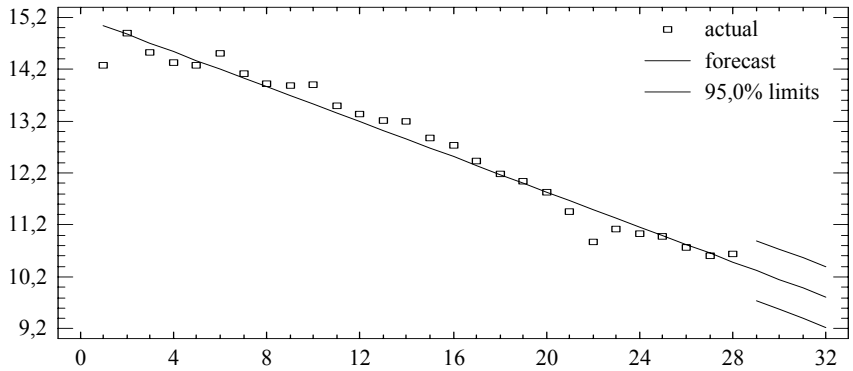
Z obr. 2.4 vidíme klesající lineární trend sezónně očištěné časové řady, jehož odhad získáme ve STAGRAPHICSu. Do vstupního panelu (obr. 2.1) vložíme do řádku časovou řadu SA_PodNez_SR, v části **Sampling Interval** označíme Other a ve Starting At necháme 1. Do políčka Number of Forecasts zapíšeme 4.

První model, který se objeví ve výsledkovém okně, je Random Walk, abychom mohli model změnit, klikneme pravým tlačítkem myši na výsledkovém okně a vybereme *Analysis Options*, objeví se doplňkový panel ***Model Specification Options*** (obr. 2.2), kde zvolíme lineární trend. Odhad parametrů lineárního trendu je v tab. 2.2a.

Tab. 2.2a. Výsledky odhadu lineárního trendu

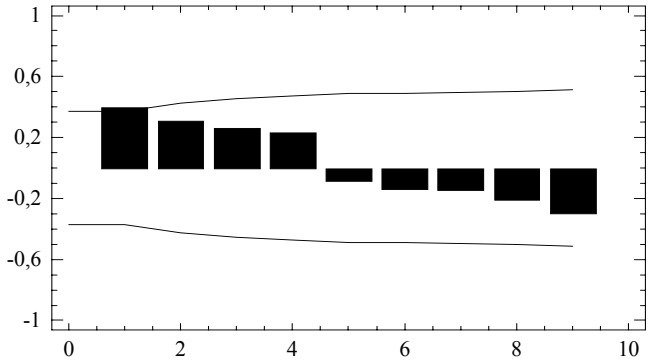
```
Data variable: SA_PodNez_SR                          Number of observations = 28
Start index = 1,0                                    Sampling interval = 1,0
Forecast model selected: Linear trend = 15,2067 - 0,168535 t

                              Trend Model Summary
Parameter      Estimate        Stnd. Error         t           P-value
----------------------------------------------------------------------
Constant       15,2067         0,101022         150,529       0,000000
Slope          -0,168535       0,00608631       -27,6909      0,000000
```

Podle tabulky 2.2a lineární trend SA_PodNez_SR$_t$ = 15,2067 – 0,168535t pro t = 1, 2, ..., 28, kde t = 1 značí hodnotu z prvního čtvrtletí roku 1994 a t = 28 značí hodnotu čtvrtého čtvrtletí roku 2000. Lineární trend má statisticky významné odhady parametrů na 5% hladině významnosti a jeho vyrovnané a extrapolované hodnoty jsou na obr. 2.4a.



Obr. 2.4a: Vyrovnané a extrapolované hodnoty lineárního trendu

Nyní ověříme vhodnost tohoto trendu. Autokorelaci nesystematické složky zjistíme grafickou analýzou reziduální autokorelační funkce, kterou získáme pomocí (obr. 2.4b).



Obr. 2.4b: Reziduální autokorelační funkce lineárního trendu

Z obr. 2.4b se zdá, že první koeficient autokorelace reziduí je na 5% hladině statisticky významný. Toto ověříme tak, že porovnáme odhady koeficientů autokorelace s intervaly spolehlivosti. Konkrétní odhady koeficientů autokorelace, odhady směrodatných chyb a 95% intervaly spolehlivosti získáme z , volbou Residual Autocorrelations (tab. 2.2b).

Tab. 2.2b. Reziduální autokorelační funkce

```
Estimated Autocorrelations for residuals
Data variable: SA_PodNez_SR
Model: Linear trend = 15,2067 -0,168535 t
                                        Lower 95,0%   Upper 95,0%
Lag  Autocorrelation  Stnd. Error    Prob. Limit   Prob. Limit
-------------------------------------------------------------
1       0,393888       0,188982       -0,370399      0,370399
2       0,308645       0,216324       -0,423989      0,423989
3       0,25905        0,231518       -0,453768      0,453768
4       0,232505       0,241649       -0,473623      0,473623
5      -0,0862542      0,24951        -0,489032      0,489032
6      -0,141524       0,250573       -0,491115      0,491115
7      -0,147548       0,253412       -0,496679      0,496679
8      -0,214985       0,256461       -0,502656      0,502656
9      -0,299199       0,262819       -0,515117      0,515117
```

Z tab. 2.2b je zřejmé, že první odhad prvního koeficientu autokorelace $r_1 = 0{,}393888$ přesahuje horní mez intervalu spolehlivosti, z čehož vyplývá, že rezidua lineárního trendu vykazují korelační závislost nesystematické složky a lineární trend je tedy nevhodný.

Korelační závislost nesystematické složky lze potvrdit i pomocí Durbinovy–Watsonovy statistiky, kterou získáme ve v proceduře **Multiple Regression** (tab. 2.2c).

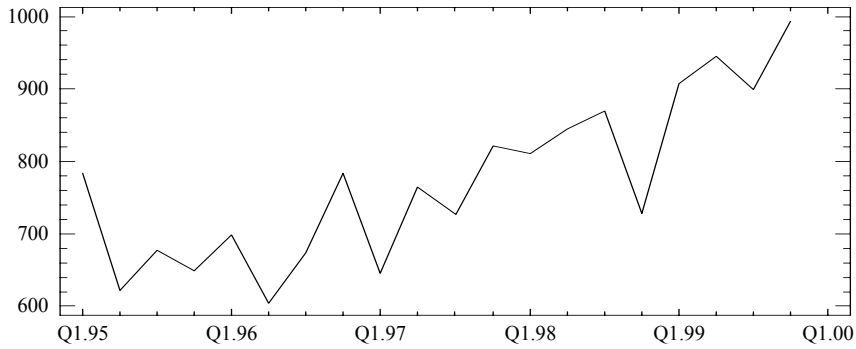
Tab. 2.2c. Výřez z výsledkového okna Multiple Regression

```
R-squared = 96,7204 percent
R-squared (adjusted for d.f.) = 96,5943 percent
Mean absolute error = 0,181972
Durbin-Watson statistic = 0,865911
```

Hodnota Durbinovy-Watsonovy statistiky $DW = 0{,}865911 < 2$, takže se jedná o pozitivní autokorelaci. Z uvedeného vyplývá, že je třeba aplikovat jiný model trendu.

Příklad 2.2

Jsou dány čtvrtletní údaje sezónně očištěné časové řady spotřeby elektrické energie firmy Stella v letech 1995 – 1999 (SA_SpEn_SR) v kWh (obr. 2.5). Cílem analýzy je najít model trendu a ověřit jeho vhodnost v období interpolace v letech 1995 - 1997, otestovat autokorelaci nesystematické složky, určit předpovědi "ex post" na roky 1998 - 1999 a posoudit zda je vybraný model vhodný na předpovídání.



Obr. 2.5: Sezónně očištěná čtvrtletní časová řada spotřeby elektrické energie firmy Stella v kWh v letech 1995 - 1999

Z obr. 2.5 se zdá, že vývoj spotřeby elektrické energie má kvadratický trend. Období analýzy rozdělíme na dvě části. Na období interpolace (prvních 12 údajů) a období verifikace modelu (posledních 8 údajů).

Ve STAGRAPHICSu vyplníme vstupní panel procedury **Forecasting** (obr. 2.1) následujícím způsobem, do řádku Data vložíme časovou řadu SA_SpEn_SR, v části **Sampling Interval** označíme Other a ve Starting At necháme 1. Do políčka Number of Forecasts zapíšeme 4 a do Withhold for Validation vepíšeme 8.

Výsledky odhadu kvadr. trendu pro období I/1995 až IV/1997 jsou v tab. 2.3a.

Tab. 2.3a. Výsledky odhadu trendu v období I/1995 až IV/1997

```
Data variable: SA_SpEn_SR                          Number of observations = 20
Start index = 1,0                                  Sampling interval = 1,0
Length of seasonality = 4
Forecast model selected: Quadratic trend = 756,294 - 36,623 t + 3,42894 t^2
Number of forecasts generated: 4
Number of periods withheld for validation: 8

                        Trend Model Summary
Parameter      Estimate   Stnd. Error      t        P-value
-----------------------------------------------------------
Constant       756,294     60,1026      12,5834     0,000001
Slope          -36,623     21,2569      -1,7229     0,119008
Quadratic        3,429      1,5918       2,1541     0,059633
```

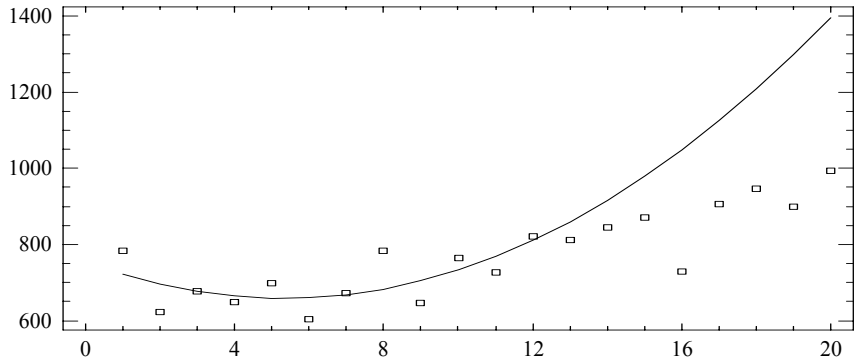
Z tab. 2.3a je zřejmé, že na 5% hladině významnosti nejsou koeficienty přírůstku a zrychlení statisticky významné. Kvadratický trend proto není vhodně zvolený.

O nevhodnosti kvadratického trendu v daném období svědčí i průměrné charakteristiky v období interpolace a extrapolace "ex post", které obsahuje tab. 2.3b.

Tab. 2.3b. Průměrné charakteristiky reziduí (Estimation Period) a chyb předpovědí "ex post" (Validation Period)

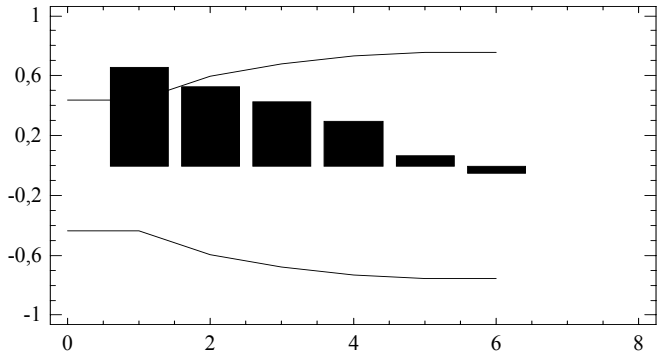
```
                  Estimation        Validation
Statistic           Period            Period
---------------------------------------------
MSE               3381,750           69752,6
MAE                41,1306           228,533
MAPE               5,90163           25,9078
ME              -1,80004E-13        -228,533
MPE             -0,521021           -25,9078
```

Z tab. 2.3b vidíme, že průměrné charakteristiky v čase extrapolace "ex post" jsou podstatně vyšší než v období interpolace, což vylučuje použití kvadratického modelu trendu na předpovídání. Podle záporných znamének ME a MPE usuzujeme, že model systematicky nadhodnocuje skutečnost, toto je zřejmé i z obr. 2.5a pro posledních osm hodnot.



Obr. 2.5a: Vyrovnané hodnoty (I/95–IV/97) a extrapolace ex post (I/98-IV/99)

Podobně z grafu reziduální autokorelační funkce zjišťujeme, že nesystematická složka kvadratického trendu je korelačně závislá na 5% hladině významnosti, protože první koeficient autokorelace překračuje horní hranici 95% intervalu spolehlivosti.



Obr. 2.5b: Reziduální autokorelační funkce kvadratického trendu.

Pokud odhadneme kvadratický trend na celé časové řadě od I/95 do IV/99, dostaneme výsledky v tab. 2.3c.

Tab. 2.3c. Odhad kvadratického trendu v období I/95 až IV/99

```
Data variable: SA_SpEn_SR                Number of observations = 20
Length of seasonality = 4
Forecast model selected: Quadratic trend = 692,38 - 7,11737 t  + 1,07836 t^2
Number of forecasts generated: 4   Number of periods withheld for validation: 0
                        Trend Model Summary
Parameter      Estimate    Stnd. Error     t         P-value
-----------------------------------------------------------
Constant       692,38      44,9765      15,3943     0,000000
Slope         -7,11737     9,86404      -0,721547   0,480380
Quadratic      1,07836     0,456258      2,36348    0,030278
```

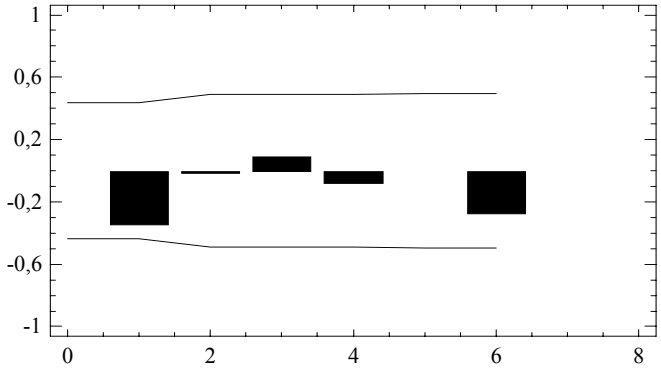
Z této tabulky je zřejmé, že koeficient přírůstku je na rozdíl od koeficientu zrychlení stále na 5% hladině významnosti statisticky nevýznamný. Upravíme proto kvadratický trend tak, že tento koeficient z modelu vyloučíme.

Ve STATGRAPHICSu k tomuto účelu využijeme proceduru **Multiple Regression**, kde ve vstupním panelu vložíme do řádku Dependent Variable časovou řadu SA_SpEn_SR a do políčka Independent Variables časovou proměnnou t^2 ve tvaru `count(1;20;1)^2`, protože obsahuje posloupnost hodnot 1, 2, ..., 20. Výsledný odhad modelu je v tab. 2.3d.

Tab. 2.3d. Upravený odhad kvadratického trendu v období I/95 až IV/99

```
Dependent variable: SA_SpEn_SR
                                          Standard          T
Parameter                Estimate          Error        Statistic        P-Value
-----------------------------------------------------------------------------
CONSTANT                  663,536         20,3358         32,629          0,0000
count(1;20;1)^2          0,758578        0,106981        7,09078          0,0000
-----------------------------------------------------------------------------
                            Analysis of Variance
-----------------------------------------------------------------------------
Source              Sum of Squares     Df  Mean Square    F-Ratio       P-Value
-----------------------------------------------------------------------------
Model                     178859,0      1     178859,0      50,28        0,0000
Residual                   64031,8     18      3557,32
-----------------------------------------------------------------------------
Total (Corr.)             242891,0     19

R-squared = 73,6376 percent
R-squared (adjusted for d.f.) = 72,173 percent
Standard Error of Est. = 59,6433
Mean absolute error = 46,0584
Durbin-Watson statistic = 2,49652
```

Analýzou reziduí na základě výběrové autokorelační funkce (obr. 2.5c) se vhodnost kvadratického trendu potvrdila, protože všechny koeficienty autokorelace jsou statisticky nevýznamné na 5% hladině významnosti.

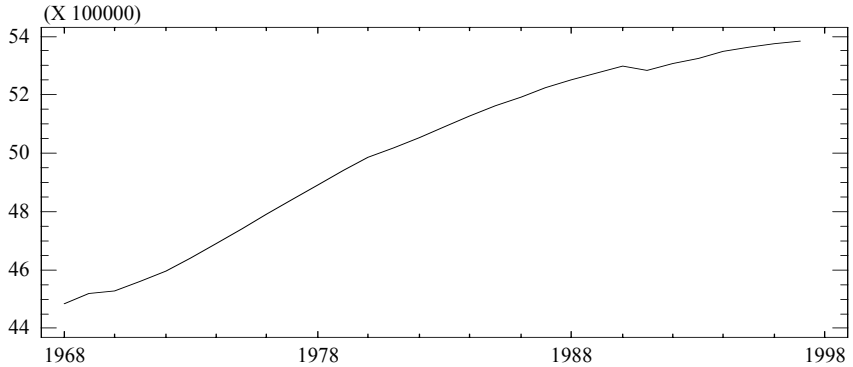


Obr. 2.5c: Reziduální autokorelační funkce kvadratického trendu v období I/95-IV/99

Pro výpočet charakteristiky MSE si ve STATGRAPHICSu uložíme rezidua pomocí a volbou položky Residuals. Hodnotu MSE si vypočítáme v panelu **Generate Data**, kde do řádku Expression vepíšeme příkaz `sum(residuals^2)/20`. V tabulkovém editoru STATGRAPHICSu takto získáme MSE = 3201,587. Z MSE lze usuzovat, že sezónně očištěné hodnoty řady spotřeby elektrické energie firmy Stella budou kolísat kolem kvadratického trendu o $\pm\sqrt{3201{,}587} = 56{,}58$ kWh.

Příklad 2.3

Je dána roční časová řada středního stavu obyvatel Slovenska (Obyv_SR) v mil. osob v letech 1968 – 1997 (obr. 2.6). Odhadněte logistický trend a ověřte autokorelaci nesystematické složky modelu, určete očekávaný počet obyvatel Slovenska v letech 1998 - 2000.



Obr. 2.6: Vývoj středního stavu obyvatel na Slovensku v letech 1968 - 1997

Z grafu vidíme, že v posledních sedmi letech se meziroční přírůstek obyvatelstva zpomalil a konverguje k určité konstantě. Volíme logistický trend ve tvaru Pearlovy–Reedovy funkce (2.24), který se používá na projekce počtu obyvatel.

Vyrovnané hodnoty a předpovědi na období 1998 - 2000 pro střední stav obyvatel Slovenska získáme pomocí procedury nelineární regrese (Marquartovy iterační procedury) aplikované na funkci

$$\frac{10000000}{\text{Obyv68_SR}} = \gamma + \beta_0 \beta_1^{\,t}, \qquad t = 1, 2, 3, ..., 30,$$

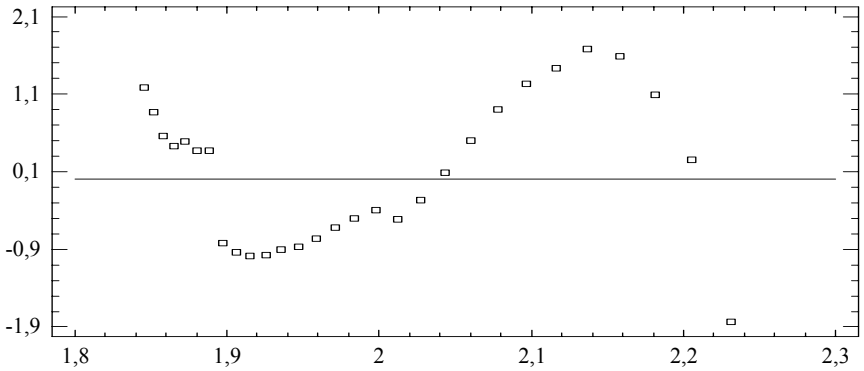
kterou jsme získali inverzní transformací vztahu (2.24).

V nelineární regresi je třeba stanovit počáteční odhady parametrů funkce, které získáme z apriorních informací nebo odhadem např. metodou tří součtů. Použijeme-li tuto metodu, získáme z údajů 10 000 000/Obyv68_SR$_t$ počáteční odhady $\hat{\gamma} = 2{,}151$, $\hat{\beta}_0 = 0{,}1416$ a $\hat{\beta}_1 = 0{,}93566$. Odhad trendu je v tab. 2.4.

Tab. 2.4: Tabulka pro odhad modifikovaného exponenciálního trendu středního stavu obyvatel SR

```
Dependent variable: 10000000/Obyv68_SR
Independent variables: t
Function to be estimated: g+b0*b1^t
Initial parameter estimates:
   g = 2,151
   b0 = 0,1416
   b1 = 0,93566
Estimation method: Marquardt
Estimation stopped due to convergence of parameter estimates.
Number of iterations: 8
Number of function calls: 40

Estimation Results
-----------------------------------------------------------------------------
                                                      Asymptotic 95,0%
                                        Asymptotic    Confidence Interval
Parameter                Estimate  Standard Error     Lower         Upper
-----------------------------------------------------------------------------
g                         1,72753       0,0306104    1,66472       1,79034
b0                       0,558467       0,0259094   0,505305      0,611629
b1                       0,949465      0,00492661   0,939357      0,959574
-----------------------------------------------------------------------------

Analysis of Variance
-----------------------------------------------------
Source              Sum of Squares     Df  Mean Square
-----------------------------------------------------
Model                       120,86      3      40,2867
Residual                0,00366086     27  0,000135588
-----------------------------------------------------
Total                      120,864     30
Total (Corr.)             0,446153     29

R-Squared = 99,1795 percent
R-Squared (adjusted for d.f.) = 99,1187 percent
Standard Error of Est. = 0,0116442
Mean absolute error = 0,00951923
Durbin-Watson statistic = 0,257419

Residual Analysis
---------------------------------
      Estimation       Validation
n     30
MSE   0,000135588
MAE   0,00951923
MAPE  0,468744
ME    0,00000160165
MPE   -0,00288998
```

Odhad modelu je v tab. 2.4. Z výstupních charakteristik odhadu modelu dále vybíráme index determinace R^2 = 99,1795 %, což znamená, že model trendu dobře vystihuje hodnoty časové řady, a směrodatnou odchylkou reziduí $\hat{\sigma}_a$ = 0,011644. Durbinův–Watsonův test DW = 0,257419 < 2 signalizuje, že na 5% hladině významnosti jde o silnou kladnou autokorelaci, což znamená, že určitá část časové řady zůstala modelem nevysvětlená a předpovědi proto mohou být zkreslené. Charakter nevysvětlené složky lze analyzovat grafem reziduí (obr. 2.6a).



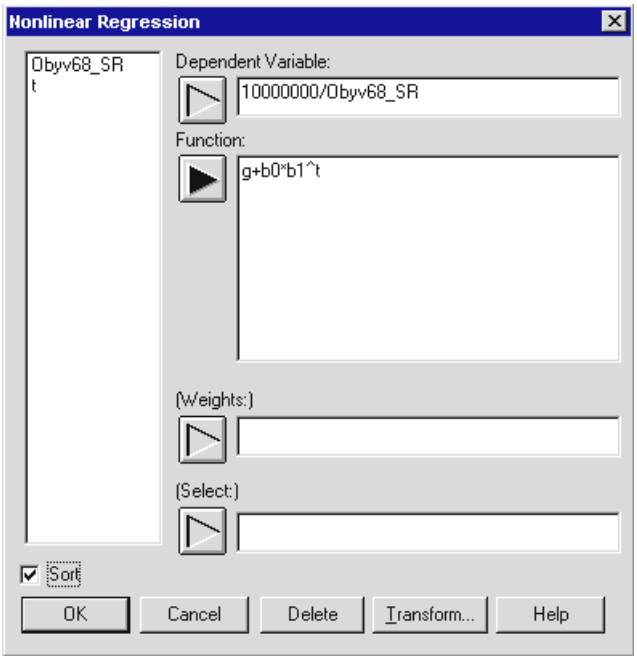
Obr. 2.6a: Graf reziduí

Z obr. 2.6a je zřejmé, že v časové řadě středního stavu obyvatelstva je kromě logistického trendu také cyklická složka, která zůstala modelem nevysvětlená. Před konstrukcí předpovědí (hlavně z dlouhodobého hlediska) je proto potřeba, buď model rozšířit, aby byla zachycena také cyklická složka časové řady, nebo zvolit jiný komplexnější model.

Pro odhad modifikovaného exponenciálního trendu ve STATGRAPHICSu vybereme z nabídek postupně

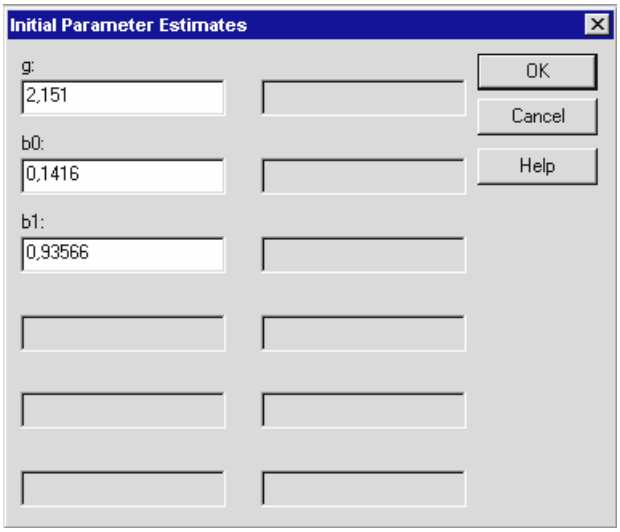
Special
⇒ **Advanced Regression**
⇒ **Nonlinear Regression**.

Vstupní panel nelineární regrese vyplníme dle obr. 2.6b, kde časová řada Obyv68_SR obsahuje zkrácenou řadu Obyv_SR od roku 1968. Časová proměnná *t* obsahuje posloupnost hodnot 1, 2, ..., 30.



Obr. 2.6b: Vyplnění vstupního panelu nelineární regrese

Obdobně podle obr. 2.6c vyplníme i panel pro zadání počátečních odhadů.



Obr. 2.6c: Vyplnění vstupního panelu pro zadání počátečních odhadů nelineární regrese

2.2.3 Klouzavé průměry

Klasická analýza časových řad předpokládá, že trendová funkce má v čase konstantní parametry. V delším časovém období je tento předpoklad nereálný (viz příklady 2.2 a 2.3), proto je vhodné využívat adaptivní techniky, jako jsou metoda klouzavých průměrů a exponenciální vyrovnávání.

Metoda klouzavých průměrů

Metoda klouzavých průměrů se zakládá na myšlence, že časovou řadu y_t pro $t = 1, 2, ..., T$ rozdělíme na kratší časové úseky o počtu hodnot $2m + 1$, na kterých odhadujeme lokální polynomické trendy určitého stupně. Pro orientaci, konstantní trend se popisuje polynomem nultého stupně, lineární trend polynomem prvního stupně apod.

Postup je následující - první část řady má $2m + 1$ hodnot, které označujeme $y_1, y_2, ..., y_{2m+1}$, z nich odhadneme parametry lokálního trendu vhodným polynomem a vypočítáme jeho odhad $\hat{T}_{m+1}$, stejný polynom odhadneme na druhé skupině hodnot řady, $y_2, y_3, ..., y_{2m+2}$ a vypočítáme odhad lokálního trendu $\hat{T}_{m+2}$, tímto klouzavým způsobem postupujeme až do konce časové řady.

Ve skutečnosti polynom lokálního trendu nemusíme odhadovat na každém úseku řady, protože se jedná o vytváření lineárních kombinací hodnot původní časové řady s pevnými koeficienty, které jsou dány zvoleným typem trendové funkce a délkou klouzavých průměrů.

V práci Kendall, M. (1976) se uvádějí koeficienty (váhy) pro různé typy polynomů a délky klouzavých částí. Některé z nich uvádíme v tabulce 2.5. Pro ilustraci nyní ukážeme způsob získávání koeficientů pro kvadratický trend.

Předpokládejme, že řadu y_t pro $t = 1, 2, ..., T$ budeme vyrovnávat na klouzavých úsecích délky $2m + 1 = 5$ hodnot parabolickým trendem s časovou proměnnou $\tau = -2, -1, 0, 1, 2$. Odhady parametrů lokálního kvadratického trendu $T_\tau = \beta_0 + \beta_1 \tau + \beta_2 \tau^2$ získáme metodou nejmenších čtverců splněním podmínky

$$\sum_{\tau=-2}^{2} (y_{t+\tau} - \beta_{0t} - \beta_{1t}\tau - \beta_{2t}\tau^2)^2 \Rightarrow \min. \tag{2.50}$$

Řešením jsou normální rovnice paraboly

$$\begin{aligned}
\sum_{\tau=-2}^{2} y_{t+\tau} &= 5\hat{\beta}_{0t} + \hat{\beta}_{1t} \sum_{\tau=-2}^{2} \tau + \hat{\beta}_{2t} \sum_{\tau=-2}^{2} \tau^2, \\
\sum_{\tau=-2}^{2} \tau y_{t+\tau} &= \hat{\beta}_{0t} \sum_{\tau=-2}^{2} \tau + \hat{\beta}_{1t} \sum_{\tau=-2}^{2} \tau^2 + \hat{\beta}_{2t} \sum_{\tau=-2}^{2} \tau^3, \\
\sum_{\tau=-2}^{2} \tau^2 y_{t+\tau} &= \hat{\beta}_{0t} \sum_{\tau=-2}^{2} \tau^2 + \hat{\beta}_{1t} \sum_{\tau=-2}^{2} \tau^3 + \hat{\beta}_{2t} \sum_{\tau=-2}^{2} \tau^4.
\end{aligned} \tag{2.51}$$

Protože pro liché i platí $\sum_{\tau=-2}^{2} \tau^i = 0$, normální rovnice paraboly se zjednoduší do tvaru

$$\begin{aligned} \sum_{\tau=-2}^{2} y_{t+\tau} &= 5\hat{\beta}_{0t} + \hat{\beta}_{2t} \sum_{\tau=-2}^{2} \tau^2, \\ \sum_{\tau=-2}^{2} \tau y_{t+\tau} &= \hat{\beta}_{1t} \sum_{\tau=-2}^{2} \tau^2, \\ \sum_{\tau=-2}^{2} \tau^2 y_{t+\tau} &= \hat{\beta}_{0t} \sum_{\tau=-2}^{2} \tau^2 + \hat{\beta}_{2t} \sum_{\tau=-2}^{2} \tau^4. \end{aligned} \tag{2.52}$$

Odhadem lokálního trendu v bodě $t = m + 1 = 3$ je interpolovaná hodnota pro $\tau = 0$, která se rovná $\hat{\beta}_{0t}$. Po získaní součtů proměnné τ v soustavě (2.52) určíme $\hat{\beta}_{0t}$ řešením rovnic

$$\begin{aligned} \sum_{\tau=-2}^{2} y_{t+\tau} &= 5\hat{\beta}_{0t} + 10\hat{\beta}_{2t}, \\ \sum_{\tau=-2}^{2} \tau^2 y_{t+\tau} &= 10\hat{\beta}_{0t} + 34\,\hat{\beta}_{2t}, \end{aligned} \tag{2.53}$$

takže

$$\hat{T}_t = \hat{\beta}_{0t} = \frac{1}{35}(17 \sum_{\tau=-2}^{2} y_{t+\tau} - 5 \sum_{\tau=-2}^{2} \tau^2 y_{t+\tau})$$

a po úpravě

$$\hat{T}_t = \hat{\beta}_{0t} = \frac{1}{35}(-3y_{t-2} + 12y_{t-1} + 17y_t + 12y_{t+1} - 3y_{t+2}). \tag{2.54}$$

Lineární kombinace hodnot časové řady (2.54) se nazývá klouzavým průměrem délky 5 a řádu 2.

Ze vztahu (2.54) vyplívá, že odhad trendu v bodě t časové řady se rovná váženému klouzavému průměru pěti hodnot s uvedenými koeficienty (váhami). Kromě toho vidíme, že součet vah se rovná jedné a že váhy jsou symetrické okolo prostřední hodnoty, takže je lze zapsat ve zkrácené formě 1/35(-3, 12, 17, ...). Stejným způsobem je možné odvodit váhy pro klouzavé průměry různých délek pro libovolný polynomický trend řádu r. Pokud je řád polynomu r sudé číslo, potom klouzavé průměry řádu r a řádu $r + 1$ mají pro stejné délky klouzavé části stejné váhy.

V tabulce 2.5 jsou uvedeny váhy klouzavých průměrů 2. a 3. řádu různých délek (Kendall, M. (1976)).

Tab. 2.5. Váhy klouzavých průměrů 2. a 3. řádu

Délka klouzavé části	Váhy
3	(0, 1, 0)
5	1/35(–3, 12, 17, ...)
7	1/21(–2, 3, 6, 7, ...)
9	1/231(–21, 14, 39, 54, 59, ...)
11	1/429(–36, 9, 44, 69, 84, 89, ...)
13	1/143(–11, 0, 9, 16, 21, 24, 25, ...)
15	1/1105(–78, –13, 42, 87, 122, 147, 162, 167, ...)
17	1/323(–21, –6, 7, 18, 27, 34, 39, 42, 43, ...)
19	1/2261(–136, –51, 24, 89, 144, 189, 224, 249, 264, 269, ...)
21	1/3059(–171, –76, 9, 84, 149, 204, 249, 284, 309, 324, 329, ...)

Kromě vážených klouzavých průměrů se využívají také jednoduché klouzavé průměry, což znamená, že lokální trendy na klouzavých úsecích $2m + 1$ hodnot jsou konstantní nebo lineární a všechny hodnoty klouzavého průměru mají stejnou váhu rovnou 1.

V sezónních časových řadách se trendová složka řady odhaduje pomocí centrovaných klouzavých průměrů, protože délka klouzavé části je sudé číslo (4 nebo 12). V takovém případě vyrovnaná hodnota trendu padne mezi dvě prostřední hodnoty dané klouzavé části časové řady, což má za následek, že nelze ve stejných obdobích porovnávat původní a vyrovnané hodnoty časové řady. Tento nedostatek se odstraní tak, že počítáme jednoduché klouzavé průměry délky 2 z řady již vypočítaných klouzavých průměrů.

Volba délky klouzavé části

V časových řadách se sezónností je délka klouzavé části určená počtem sezón. V ostatních časových řadách musíme délku klouzavé části volit subjektivně, což nemusí být vždy jednoduché. Platí pravidlo, že čím hladší vyrovnání časové řady požadujeme, tím delší klouzavou část volíme.

Počáteční a koncové klouzavé průměry

Výsledkem použití metody klouzavých průměrů je odhad trendové případně trend-cyklické složky časové řady. Její nevýhodou je, že prvních a posledních m hodnot trendové nebo trend-cyklické složky není určeno. Způsob výpočtu těchto počátečních a koncových hodnot je uveden např. v pracích Kendall (1976), Rublíková (1989), Kozák, Hindls, Arlt (1994).

Ztráta posledních m klouzavých průměrů je nevýhodou zvlášť tehdy, pokud je chceme využít k prognostickým účelům. Proto byla vypracována adaptivní technika koncových předpovědních průměrů.

Předpovědní klouzavé průměry

Jednoduché klouzavé průměry je možné využít pro předpovídání časové řady s přibližně konstantním trendem (2.5). Odhad parametru β_0 se v čase t určuje jako aritmetický průměr hodnot y_t pro $t = 1, 2, ..., T$.

Předpovědi řady y_t v čase $t = T + 1, T + 2, ..., T + h$ určujeme dvěma způsoby:

1. Mají-li všechna pozorování řady stejnou váhu, odhad $\hat{\beta}_0 = \bar{y} = 1/T \sum_{t=1}^{T} y_t$ slouží jako předpověď pro libovolný horizont h s počátkem předpovídání v čase $t = T$, tj.
$$\hat{y}_T(h) = \bar{y} \qquad h > 0. \qquad (2.55)$$
2. Předpokládáme-li, že větší váhu mají pozorování bližší k předpovídané hodnotě, potom předpověď určujeme jako průměr z posledních n hodnot $y_T, y_{T-1}, ..., y_{T-n+1}$ se stejnou váhou $1/n$ a ostatním hodnotám řady $y_{T-n}, y_{T-n-1}, ..., y_1$ přiřadíme nulovou váhu. Odhad parametru modelu β_0 získáme metodou nejmenších čtverců, jako průměr z posledních n hodnot řady

$$\hat{\beta}_0 = \frac{1}{n} \sum_{t=T-n+1}^{n} y_t = MA_T ,$$

resp.

$$MA_T = \frac{y_T + y_{T-1} + y_{T-2} + \ldots + y_{T-n+1}}{n} . \tag{2.56}$$

Předpověď časové řady od období $t = T$ pro horizont $h = 1$ je

$$\hat{y}_T(h) = \hat{\beta}_0 = MA_T . \tag{2.57}$$

Předpovědi "ex post" je možné určovat od začátku časové řady a tak ověřovat jejich kvalitu. Předpověď v čase t označujeme symbolem $\hat{y}_{t-1}(t)$, tzn. že předpověď určujeme v čase t - 1 s horizontem $h = 1$ (tedy na čas t).

Poznámka:

Předpovědi se určují s horizontem $h = 1$. Pro vzdálenější horizonty předpověď $\hat{y}_T(h) = MA_T$ nereaguje na skutečnost, že jsme mezitím získali nové pozorování y_{T+1}, které vytvoří novou skupinu posledních n hodnot (nejstarší pozorování se z řady vyloučí a přidá se nové). Nový klouzavý průměr může být jiný než $\hat{y}_T(T+2) = MA_T$.

2.2.4 Klouzavé průměry ve STATGRAPHICSu

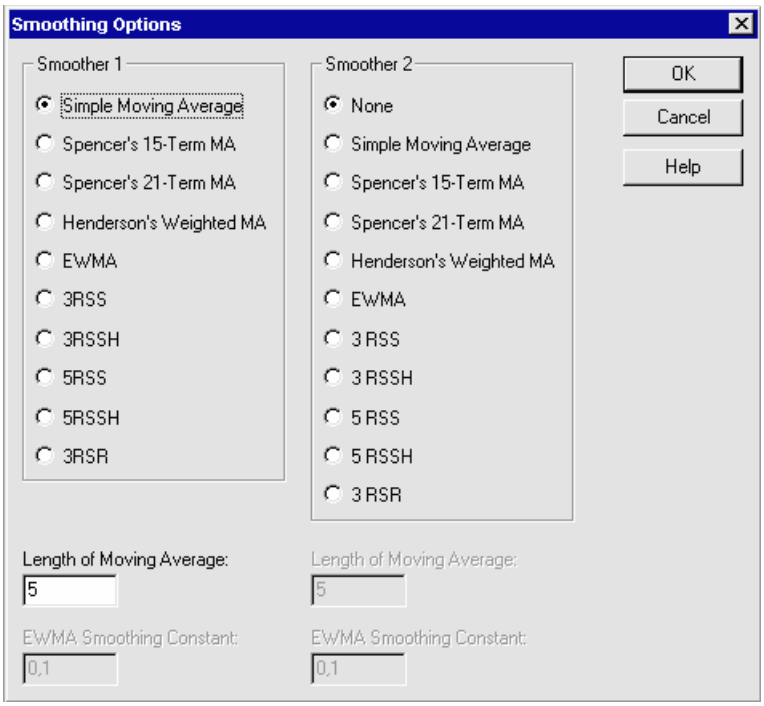
Podle toho, na co klouzavé průměry používáme, zda pro vyrovnávání (odhad trendu) nebo pro předpovídání (předpovědi trendu), můžeme ve STATGRAPHICSu použít dvě procedury: **Smoothing** a **Forecasting.**

Odhady trendu pomocí klouzavých průměrů různých délek získáme v proceduře **Smoothing**

Special
⇒ **Time-Series Analysis**
⇒ **Smoothing**.

Po vyplnění vstupního panelu vybereme z položku Data Table, kde se automaticky objeví tabulka klouzavých průměrů s délkou klouzavé části 5 hodnot a označuje se jako *Simple moving averages of 5 terms.*

Délku klouzavé části změníme kliknutím pravého tlačítka myši a volbou *Pane Options* z doplňkové nabídky. Zobrazí se doplňkový panel ***Smoothing Options*** (obr. 2.8).



Obr. 2.8: Doplňkový panel Smoothing

Délku klouzavé části (Lenght of Moving Average) zadáváme počtem $2m+1$ hodnot, z kterých určujeme klouzavý průměr.

Grafy vyrovnaných hodnot trendu získáme z , kde vybereme Time Sequence Plot a Residual Plot.

Z obr. 2.8 vidíme, že kromě jednoduchých klouzavých průměrů lze vybírat i různé vážené klouzavé průměry, jako jsou např. Spencerovy či Hendersonovy klouzavé průměry (viz Kendall (1976)).

Předpovědní klouzavé průměry ve STATGRAPHICSu

Jednoduché klouzavé průměry slouží nejen k odhadu trendové složky v období interpolace, ale využívají se i pro předpovídání časových řad s konstantním trendem. Ve STATGRAPHICSu je najdeme v proceduře **Forecasting**

Special
⇒ **Time-Series Analysis**
⇒ **Forecasting.**

V doplňkovém panelu ***Model Specification Options*** (obr. 2.2) vybereme v části **Type** položku Moving Average a současně v části **Parameters and Terms** položku Order, kam zadáváme délku klouzavé části rovnou posledním n hodnotám, ze kterých určujeme klouzavý průměr. Pro zobrazení výsledků vybíráme z položku Forecast Table. Grafy získáme z volbou Time Sequence Plot.

Příklad 2.4

Jsou dány údaje o počtu živě narozených dětí v České republice v letech 1920 - 1999 (ZivNar_CR). Vyrovnejte časovou řadu jednoduchými klouzavými průměry s délkou klouzavé části 5 let. Graficky zobrazte původní údaje a jejich vyrovnané hodnoty (jednoduché klouzavé průměry) pro délky klouzavé části 5 a 15 let a pozorujte, jak volba délky klouzavé části ovlivňuje vyrovnání časové řady.

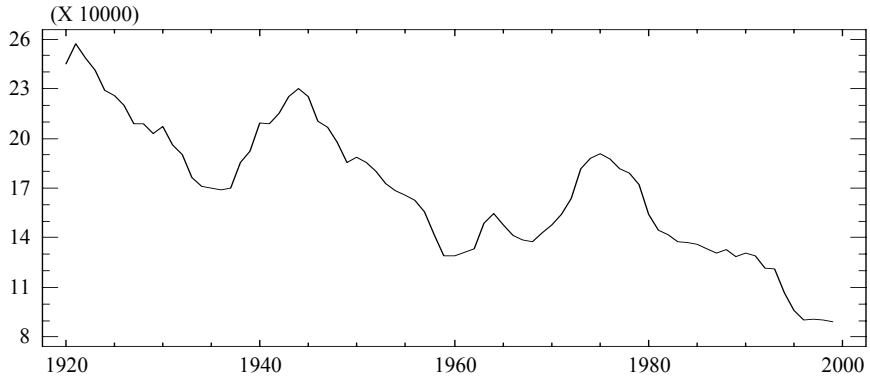
Z obr. 2.9 vidíme, že počet živě narozených dětí v České republice má ve sledovaném období klesající trend a výraznou cyklickou složku. Výsledky vyrovnávání časové řady klouzavými průměry různých délek lze pozorovat na obr. 2.9a a obr. 2.9b.

Zkrácené výsledky vyrovnávání řady klouzavými průměry s délkou klouzavé části 5 let jsou v tab. 2.6.

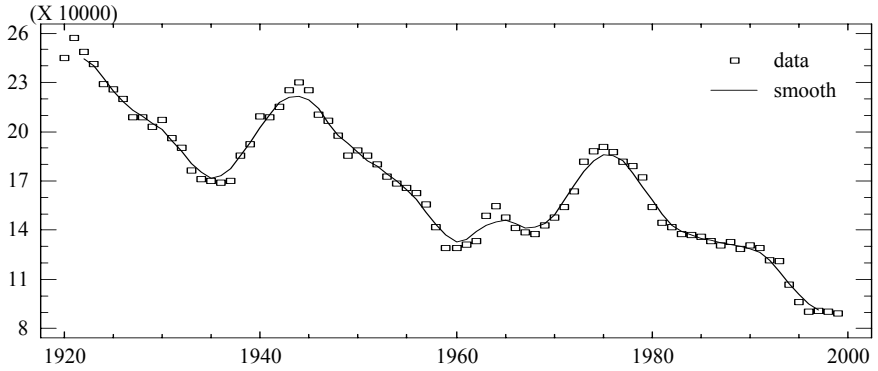
Tab. 2.6. Výsledky odhadu trendu klouzavými průměry s délkou 5 let

Data Table for ZivNar_CR
First smoother: simple moving average of 5 terms
Second smoother: none

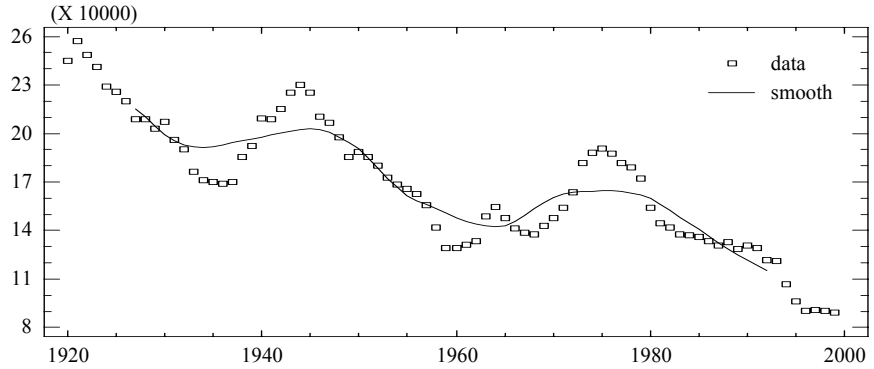
Period	Data	Smooth	Rough (residual)
1920	244668,0		
1921	257281,0		
1922	248728,0	244160,0	4567,8
1923	241230,0	240338,0	892,4
1924	228894,0	232842,0	-3947,8
1925	225555,0	224838,0	716,6
1926	219802,0	218381,0	1421,2
1927	208711,0	213215,0	-4503,8
1928	208942,0	209549,0	-606,6
1929	203064,0	204831,0	-1767,0
1930	207224,0	201168,0	6055,8
.			
.			
1983	137431,0	139286,0	-1854,8
1984	136941,0	137069,0	-128,4
1985	135881,0	134906,0	975,0
1986	133356,0	133953,0	-597,2
1987	130921,0	132236,0	-1315,2
1988	132667,0	131173,0	1494,2
1989	128356,0	130372,0	-2016,4
1990	130564,0	128529,0	2034,8
1991	129354,0	126201,0	3153,2
1992	121705,0	121845,0	-140,4
1993	121025,0	114952,0	6073,0
1994	106579,0	107170,0	-591,4
1995	96097,0	100961,0	-4863,8
1996	90446,0	94862,8	-4416,8
1997	90657,0	91441,2	-784,2
1998	90535,0		
1999	89471,0		



Obr. 2.9: Vývoj počtu živě narozených dětí v ČR



Obr. 2.9a: Klouzavé průměry délky 5

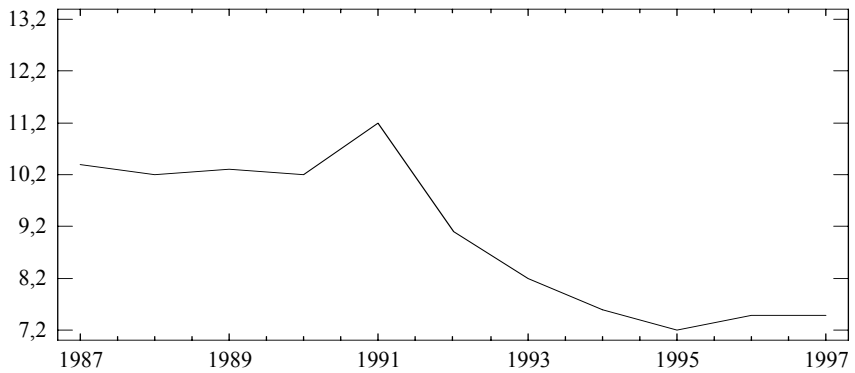


Obr. 2.9b: Klouzavé průměry délky 15

Na obr. 2.9a vidíme, že pokud je délka klouzavé části malá (5 hodnot), vyrovnání je podobné původní časové řadě na obr. 2.9. Při vyrovnávání klouzavými průměry s délkou 15 hodnot (obr. 2.2.9b) je trendová čára hladší. Čím je délka klouzavé části větší, tím větší je i vyrovnání (vyhlazení) časové řady a naopak.

Příklad 2.5

Jsou dány údaje o podílu pracujících ve stavebnictví na Slovensku (PracStav_SR) v letech 1987 - 1997. Analyzujte vývoj podílu pracujících ve stavebnictví za uvedené období a pomocí klouzavých průměrů odhadněte tento podíl pro rok 1998.



Obr. 2.10: Podíl pracujících ve stavebnictví na Slovensku (%)

Z obr. 2.10 vidíme, že v letech 1987 - 1990 byl podíl pracujících konstantní, v roce 1991 se krátkodobě zvýšil, v letech 1992 - 1993 klesal a v posledních čtyřech letech 1994 - 1997 se ustálil na 7,5 % úrovni. Předpověď odvodíme z posledních 4 let za předpokladu, že se tento vývoj nezmění a nedojde ke kvalitativním změnám. Předpověď na rok 1998 určíme jako průměr z posledních 4 hodnot časové řady.

Výpočet předpovědí je velmi jednoduchý a zdá se, že nepotřebujeme počítač. Počítač však pomáhá vyhodnotit nejen předpovědní techniku v době interpolace, ale i průměrné charakteristiky chyb předpovědí s horizontem $h = 1$.

Výsledky získané ve STATGRAPHICSu obsahuje tab. 2.7a) a 2.7b) spolu s 95 % intervalem spolehlivosti pro skutečnou hodnotu podílu pracujících ve stavebnictví v roce 1998.

Tab. 2.7a. Průměrné charakteristiky chyb předpovědí "ex post" s $h = 1$

```
Data variable: PracStav_SR                Number of observations = 11
Start index = 1987
Forecast model selected: Simple moving average of 4 terms
Number of forecasts generated: 2          Number of periods withheld for validation: 0

             Estimation      Validation
Statistic     Period          Period
--------------------------------------------
MSE          2,09625
MAE          1,26429
MAPE         15,58220
ME          -1,0000
MPE         -13,2225
```

Tab. 2.7b. Výsledky předpovědí "ex post" a "ex ante" s $h = 1$

```
Forecast Table for PracStav_SR
Model: Simple moving average of 4 terms

Period           Data               Forecast             Residual
-----------------------------------------------------------------
1987             10,4
1988             10,2
1989             10,3
1990             10,2
1991             11,2               10,275                0,925
1992             9,1                10,475               -1,375
1993             8,2                10,200               -2,0
1994             7,6                9,675                -2,075
1995             7,2                9,025                -1,825
1996             7,5                8,025                -0,525
1997             7,5                7,625                -0,125

                                     Lower 95.0%         Upper 95.0%
Period           Forecast            Limit               Limit
-----------------------------------------------------------------
1998             7,45                4,27733             10,6227
```

2.2.5 Exponenciální vyrovnávání

Mějme časovou řadu y_t. Jednoduché exponenciální vyrovnávání časové řady je definované rekurentním vztahem

$$\hat{y}_t = \alpha y_t + (1 - \alpha)\, \hat{y}_{t-1}, \tag{2.55}$$

kde $\hat{y}_t$ je exponenciální průměr v čase t, $\hat{y}_{t-1}$ je exponenciální průměr v čase t - 1 a α je vyrovnávací konstanta, $\alpha \in \langle 0, 1 \rangle$.

Na základě rekurentnosti lze exponenciální průměr vyjádřit následovně

$$\begin{aligned}
\hat{y}_t &= \alpha y_t + (1-\alpha)\hat{y}_{t-1} = \alpha y_t + (1-\alpha)\left[\alpha y_{t-1} + (1-\alpha)\hat{y}_{t-2}\right] = \\
&= \alpha y_t + \alpha(1-\alpha) y_{t-1} + (1-\alpha)^2\left[\alpha y_{t-2}(1-\alpha)\hat{y}_{t-3}\right] = \ldots = \\
&= \alpha y_t + \alpha(1-\alpha) y_{t-1} + (1-\alpha)^2 \alpha y_{t-2} + \ldots + \alpha(1-\alpha)^i y_{t-i} + \ldots + (1-\alpha)^t \hat{y}_0 = \\
&= \alpha \sum_{i=0}^{t-1} (1-\alpha)^i y_{t-i} + (1-\alpha)^t \hat{y}_0
\end{aligned} \tag{2.56}$$

kde $\hat{y}_0$ je počáteční podmínka pro výpočet vztahu (2.55).

Za předpokladu, že (1 - α) < 1, pro $t \to \propto$ platí (1 - α)$^t \to 0$, takže vztah (2.56) má tvar

$$\hat{y}_t = \alpha \sum_{i=0}^{\infty} (1-\alpha)^i y_{t-i},$$

váženého součtu všech hodnot časové řady s vahami, které směrem do minulosti exponenciálně klesají. To je důvod, proč se $\hat{y}_t$ nazývá exponenciálním průměrem v čase t.

Brownovo jednoduché exponenciální vyrovnávání

Předpokládejme, že časová řada y_t je generována stacionárním procesem (viz kapitola 3) s modelem ve tvaru

$$y_t = \beta_0 + a_t, \tag{2.57}$$

kde β_0 je střední hodnota procesu, a_t jsou pro každé t náhodné veličiny s vlastnostmi procesu bílého šumu (2.4).

Po aplikaci metody exponenciálního vyrovnávání na řadu (2.57) dostaneme

$$\hat{y}_t = \alpha \sum_{i=0}^{\infty} (1-\alpha)^i y_{t-i} = \alpha \sum_{i=0}^{\infty} (1-\alpha)^i (\beta_0 + a_{t-i}) = \beta_0 + \alpha \sum_{i=0}^{\infty} (1-\alpha)^i a_{t-i}, \tag{2.58}$$

neboť $\alpha \sum_{i=1}^{\infty} (1-\alpha) = 1$.

Lze dokázat, že pro střední hodnotu a rozptyl vztahu (2.58) platí

$$E(\hat{y}_t) = E(y_t) = \beta_0,$$

$$D(\hat{y}_t) = \frac{\alpha}{2-\alpha}\sigma_a^{\ 2} = \frac{\alpha}{2-\alpha}\sigma_y^{\ 2}. \tag{2.59}$$

Ze vztahů (2.59) vyplívá, že pro $\alpha \in \langle 0, 1\rangle$, je exponenciální průměr $\hat{y}_t$ v čase t nezkresleným odhadem parametru β_0 (tj. platí $\hat{y}_t = \hat{\beta}_0$) a má menší rozptyl než je rozptyl časové řady y_t. Tato vlastnost vysvětluje funkci vyrovnávací konstanty α.

Pokud je α blízká jedné, potom $D(\hat{y}_t) \cong D(y_t) = \sigma_y^2$. To znamená, že se z řady nepodařilo vyloučit nesystematickou složku a vyrovnání časové řady je nedostatečné. Naopak, pokud je α blízká nule, je rozptyl exponenciálních průměrů v porovnaní s rozptylem časové řady menší. To znamená, že pro malé α se nesystematická složka z řady více eliminuje a vyrovnání řady je zřetelnější.

Vlivem publikací Brown (1959), (1961) se exponenciální vyrovnávání začalo využívat jako jednoduchá technika krátkodobého předpovídání s horizontem $h = 1$ období dopředu.

Předpovídání pomocí exponenciálního vyrovnávání předpokládá, že časovou řadu vyjádříme modelem

$$y_t = \beta_{0,t} + a_t, \tag{2.60}$$

kde $\beta_{0,t}$ je podmíněná střední hodnota časové řady, měnící se v závislosti na t a a_t jsou pro každé t náhodné veličiny s vlastnostmi (2.4).

Odhadem parametru $\beta_{0,t}$ v čase t je $\hat{y}_t$, takže platí $\hat{\beta}_{0,t} = \hat{y}_t$.

Bodová předpověď a její chyba

Mějme časovou řadu $y_1, y_2, ..., y_{t-1}, y_t$. Bodovou předpověď v čase $t + 1$, definujeme jako exponenciální průměr řady v čase t a označujeme

$$\hat{y}_t(1) = \hat{\beta}_{0,\mathrm{t}} = \hat{y}_t, \tag{2.61}$$

resp.

$$\hat{y}_{t-1}(1) = \hat{\beta}_{0,t-1} = \hat{y}_{t-1}.$$

Chybou předpovědi je rozdíl skutečné a předpovídané hodnoty v čase t,

$$\hat{a}_t = y_t - \hat{y}_{t-1}(1). \tag{2.62}$$

<u>Poznámka:</u>

Výraz $\hat{y}_{t-1}(1)$ znamená bodovou předpověď určovanou v čase t - 1 s horizontem 1 období dopředu, tedy na čas t.

Předpovídání formou korekce chyby předpovědi

Upravíme-li vztah (2.55) do tvaru

$$\hat{y}_t = \alpha y_t + (1-\alpha)\hat{y}_{t-1} = \alpha y_t + \hat{y}_{t-1} - \alpha\hat{y}_{t-1} = \alpha(y_t - \hat{y}_{t-1}) + \hat{y}_{t-1} = \alpha\hat{a}_t + \hat{y}_{t-1},$$

a uplatníme-li vztahy (2.61) a (2.62) zjistíme, že

$$\hat{y}_t(1) = \alpha\hat{a}_t + \hat{y}_{t-1}(1), \tag{2.63}$$

tj. předpověď na čas $t + 1$ určíme jako součet předpovědi na čas t a α–násobku chyby předpovědi na čas t.

Tento přístup k určovaní předpovědí je velmi jednoduchý a pokud známe vyrovnávací konstantu a předpověď z předcházejícího období, nepotřebujeme znát žádné předcházející údaje časové řady, ale pouze hodnotu konkrétního ukazatele v čase t.

Odhad β_{0t} získáme metodou nejmenších čtverců, při splnění podmínky

$$\sum_{i=0}^{\infty}(y_{t-i} - \hat{\beta}_{0t})^2(1-\alpha)^i \Rightarrow \min. \tag{2.64}$$

95% interval spolehlivosti pro y_{T+1}

Podle Bowermana a O´Connella (1979) 100x(1 - α/2)% interval spolehlivosti pro y_{t+1} určíme vztahem

$$\hat{y}_T(1) \pm u_{1-\alpha/2} 1{,}25 MAE_T, \tag{2.65}$$

kde $\hat{y}_T(1) = \alpha y_T + (1-\alpha)\hat{\beta}_{0,T-1}$ je předpověď určená v čase $t = T$ s horizontem $h = 1$, $u_{1-\alpha/2}$ je (1 - α/2)% kvantil normovaného normálního rozdělení $N(0,1)$ a

$$MAE_T = \frac{\sum_{t=1}^{T}\left|y_t - \hat{y}_{t-1}(1)\right|}{T}.$$

Hodnotu MAE_T lze s každou novou hodnotou obnovovat, podle vztahu

$$MAE_{T+1} = \frac{T \cdot MAE_T + \left|y_{T+1} - \hat{\beta}_{0,T}\right|}{T+1}.$$

Brownovo (dvojité) lineární exponenciální vyrovnávání

Brownovo dvojité exponenciální vyrovnávání s vyrovnávací konstantou α používáme pro předpovídání časové řady y_t bez sezónní složky, jejíž vývoj lze rozložit na lokální lineární trendy vyjádřené modelem

$$y_t = T_t + a_t = \beta_0 + \beta_1 t + a_t. \tag{2.66}$$

Odhady parametrů lineárního trendu určil Brown (1959) na základě následujících úvah. Uplatnil jednoduché exponenciální vyrovnávání

$$\hat{y}_t = \alpha y_t + (1 - \alpha)\,\hat{y}_{t-1} \tag{2.67}$$

na časovou řadu s lokálně se měnícím lineárním trendem a zjistil, že řada prvních exponenciálních průměrů $\hat{y}_t$ má systematickou chybu rovnou $\beta_1((1 - \alpha)/\alpha)$, tj. očekávaná střední hodnota exponenciálních průměrů zaostává za očekávanou střední hodnotou určitého ukazatele, takže předpovědi tohoto ukazatele s horizontem $h = 1$ jsou systematicky podhodnocené. Po opětovné aplikaci jednoduchého exponenciálního vyrovnávání na časovou řadu prvních exponenciálních průměrů $\hat{y}_t$ získal řadu exponenciálních průměrů druhého stupně

$$\hat{y}_t^{(2)} = \alpha y_t + (1 - \alpha)\,\hat{y}_{t-1}^{(2)} \tag{2.68}$$

a zjistil, že se opět dopustil stejné systematické chyby. Na základě této vlastnosti určil vztahy pro výpočet odhadů $\hat{\beta}_{0,t}$ a $\hat{\beta}_{1,t}$ parametrů modelu (2.66) v čase t ve tvaru

$$\begin{aligned} \hat{\beta}_{0,t} &= 2\hat{y}_t - \hat{y}_t^{(2)}, \\ \hat{\beta}_{1,t} &= \frac{\alpha}{1-\alpha}(\hat{y}_t - \hat{y}_t^{(2)}), \end{aligned} \tag{2.69}$$

kde $\hat{\beta}_{0,t}$ je odhad parametru β_0 v čase t a $\hat{\beta}_{1,t}$ odhad parametru β_1 v čase t, $\alpha \in \langle 0, 1\rangle$ je vyrovnávací konstanta, $\hat{y}_t$ je jednoduchý exponenciální průměr získaný z řady y_t, $\hat{y}_t^{(2)}$ je dvojitý exponenciální průměr (2.68) získaný z řady exponenciálních průměrů $\hat{y}_t$.

<u>Poznámka:</u>

Pro tuto vlastnost se exponenciální vyrovnávání řad s lineární trendem nazývá dvojité exponenciální vyrovnávání.

Stejně jako při jednoduchém exponenciálním vyrovnávání, také v případě lokálně se měnících lineárních trendů, odhady parametrů modelu (2.66) můžeme získat váženou metodou nejmenších čtverců s exponenciálně klesajícími vahami (1 - α) za podmínky

$$\sum_{i=0}^{\infty}(y_t - \hat{\beta}_0 - \hat{\beta}_1 t)^2 (1-\alpha)^i \Rightarrow \min.$$

Bodová předpověď a její chyba

Bodovou předpověď určitého ukazatele pro horizont $h > 0$ od času T definujeme vztahem

$$\hat{y}_T(h) = \hat{\beta}_{0,T} + h\hat{\beta}_{1,T} = 2\hat{y}_T - \hat{y}_T^{(2)} + h\frac{\alpha}{1-\alpha}(\hat{y}_T - \hat{y}_T^{(2)}). \tag{2.70}$$

Předpověď s horizontem $h = 1$ od času (t - 1) je potom

$$\hat{y}_{t-1}(1) = \hat{\beta}_{0,t-1} + \hat{\beta}_{1,t-1}. \tag{2.71}$$

Chyba předpovědi s horizontem $h = 1$ v čase t je dána jako rozdíl skutečné hodnoty v čase t a předpovědi v čase t, tj. $\hat{a}_t = y_t - \hat{y}_{t-1}(1)$.

Dvojité exponenciální vyrovnávání ve formě korekce chyby předpovědi

Pokud známe vyrovnávací konstantu α, předpověď $\hat{y}_{t-1}(1)$ v čase t, chybu předpovědi $\hat{a}_t = y_t - \hat{y}_{t-1}(1) = y_t - \hat{\beta}_{0,t-1} - \hat{\beta}_{1,t-1}$ v čase t, potom odhady parametrů lokálně se měnícího lineárního trendu (2.66) je možné získat formou korekce chyb předpovědí $\hat{a}_t$, podle vztahů

$$\begin{aligned} \hat{\beta}_{0,t} &= \hat{y}_{t-1}(1) + \left[1-(1-\alpha)^2\right]\hat{a}_t = \hat{\beta}_{0,t-1} + \hat{\beta}_{1,t-1} + \left[1-(1-\alpha)^2\right]\hat{a}_t, \\ \hat{\beta}_{1,t} &= \hat{\beta}_{1,t-1} + \alpha^2\hat{a}_t. \end{aligned} \tag{2.72}$$

95% interval spolehlivosti pro y_{T+h}

Podle Bowermana a O´Connella (1979) 100(1 - α/2)% interval spolehlivosti pro y_{T+h} určíme vztahem

$$\hat{y}_T(h) \pm u_{1-\alpha/2} \cdot d_h \cdot MAE_T, \tag{2.73}$$

kde $\hat{y}_T(h) = \hat{\beta}_{0,T} + h\hat{\beta}_{1,T}$ je bodová předpověď určená v čase $t = T$ na čas $t = T + h$, $u_{1-\alpha/2}$ je (1 - α/2)% kvantil normovaného normálního rozdělení $N(0,1)$ a

$$d_h = 1{,}25\sqrt{\frac{1+\frac{1-\alpha}{(1+\alpha)^3}((1+4\alpha+5\alpha^2)+2(1-\alpha)(1+3\alpha)h+2(1-\alpha)^2h^2)}{1+\frac{1-\alpha}{(1+\alpha)^3}((1+4\alpha+5\alpha^2)+2(1-\alpha)(1+3\alpha)+2(1-\alpha)^2)}},$$

$$MAE_T = \sum_{t=1}^{T}\frac{\left|y_t - \hat{y}_{t-1}(1)\right|}{T}.$$

Hodnotu MAE_T je možné s každou novou hodnotou obnovovat podle vztahu

$$MAE_{T+1} = \frac{T \cdot MAE_T + \left|y_{T+1} - \hat{\beta}_{0,T} - \hat{\beta}_{1,T}\right|}{T+1}.$$

Poznámka:

Podobným způsobem lze odvodit také Brownovo trojité exponenciální vyrovnávání, které předpokládá, že časová řada má kvadratický trend (2.13).

Holtovo lineární exponenciální vyrovnávání

Holt (1957) vypracoval algoritmus exponenciálního vyrovnávání lokálních lineárních trendů (2.66) v časové řadě se dvěma vyrovnávacími konstantami α pro adaptivní odhad úrovně β_0 v čase t a β pro adaptivní odhad směrnice lineárního trendu β_1 v čase t.

Holtův algoritmus exponenciálního vyrovnávání odhaduje v čase t parametry modelu (2.66) podle rekurentních vztahů

$$\begin{aligned}\hat{\beta}_{0,t} &= \alpha y_t + (1-\alpha)(\hat{\beta}_{0,t-1} + \hat{\beta}_{1,t-1}),\\ \hat{\beta}_{1,t} &= \beta(\hat{\beta}_{0,t} - \hat{\beta}_{0,t-1}) + (1-\beta)\hat{\beta}_{1,t-1},\end{aligned} \tag{2.74}$$

kde $\hat{\beta}_{0,t}$ je odhad úrovně lineárního trendu v čase t, $\hat{\beta}_{1,t}$ je odhad směrnice lineárního trendu v čase t, $\hat{\beta}_{0,t-1}$ je odhad úrovně lineárního trendu v čase t - 1, $\hat{\beta}_{1,t-1}$ je odhad směrnice lineárního trendu v čase t - 1, $\alpha \in \langle 0, 1\rangle$ je vyrovnávací konstanta úrovně, $\beta \in \langle 0, 1\rangle$ je vyrovnávací konstanta směrnice.

Bodová předpověď a její chyba

Bodovou předpověď pro horizont $h > 0$ konstruovanou v čase t, definujeme vztahem

$$\hat{y}_t(h) = \hat{\beta}_{0,t} + h\hat{\beta}_{1,t}\,. \tag{2.75}$$

Pro $h = 1$ v čase t - 1 definujeme bodovou předpověď a její chybu vztahy

$$\hat{y}_{t-1}(1) = \hat{\beta}_{0,t-1} + \hat{\beta}_{1,t-1}\,,$$

$$\hat{a}_t = y_t - \hat{\beta}_{0,t-1} - \hat{\beta}_{1,t-1}\,.$$

Holtův model ve formě korekce chyby předpovědi

$$\begin{aligned} \hat{\beta}_{0,t} &= \hat{\beta}_{0,t-1} + \hat{\beta}_{1,t-1} + \alpha\hat{a}_t\,, \\ \hat{\beta}_{1,t} &= \hat{\beta}_{1,t-1} + \alpha\beta\hat{a}_t\,. \end{aligned} \tag{2.76}$$

Modifikace odhadů úrovně a směrnice řady závisí na volbě vyrovnávacích konstant α a β. Malé hodnoty α a β volíme, pokud se vyžadují malé modifikace odhadnuté úrovně a směrnice lineárního trendu z předcházejícího období. Softwarové produkty pracující pod operačním systémem Windows, jako jsou SAS, SPSS nebo STATGRAPHICS Plus, automaticky vyhledávají nejvhodnější kombinaci vyrovnávacích konstant α a β.

2.2.6 Exponenciální vyrovnávání ve STATGRAPHICSu

Krátkodobé předpovědi trendu pomocí exponenciálního vyrovnávání získáme v proceduře **Forecasting** výběrem některé z metod exponenciálního vyrovnávání v části **Type** v doplňkovém panelu ***Model Specifications Options*** (obr. 2.2):

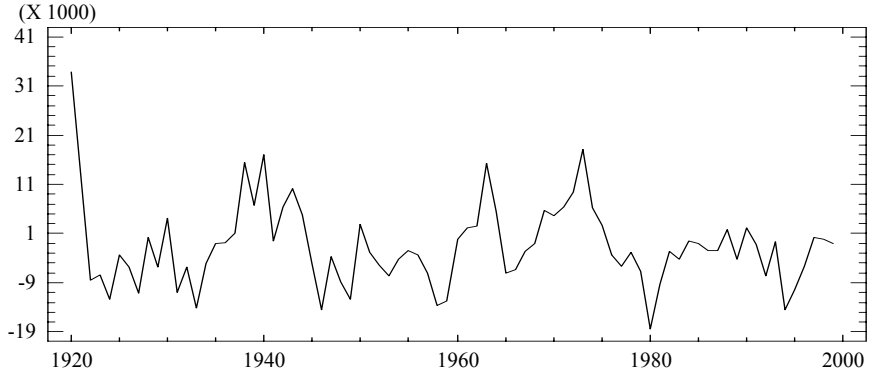
Simple Exp. Smoothing	- Brownovo jednoduché exponenciální vyrovnávání,
Brown's Linear Exp. Smoothing	- Brownovo dvojité exponenciální vyrovnávání,
Holt's Linear Exp. Smoothing	- Holtovo lineární exponenciální vyrovnávání,
Quadratic Exp. Smoothing	- Brownovo trojité exponenciální vyrovnávání,
Winter's Exp. Smoothing	- Wintersovo exponenciální vyrovnávání (pro sez. č.ř.).

V části **Parameters and Terms** označíme políčko Optimize pokud potřebujeme, aby počítač vybral nejvhodnější vyrovnávací konstantu α ve smyslu minimální hodnoty MSE (střední čtvercové chyby předpovědí s horizontem jednoho období dopředu) u Brownova exp. vyrovnávání, resp. konstanty α a β pro případ Holtova lineárního exp. vyrovnávání. Pokud chceme sami volit hodnoty těchto konstant políčko Optimize odškrtneme.

Příklad 2.6

Uvažujte časovou řadu přírůstků (úbytků) počtu živě narozených dětí v České republice v letech 1921 až 1999, jejíž vývoj je zobrazený na obr. 2.11. Tato řada se získá diferencováním původní časové řady.

Analyzujte vývoj přírůstků (úbytků) počtu živě narozených dětí v období let 1921 až 1999 v České republice. Pro uvedené období navrhněte vhodný model exponenciálního vyrovnávání a interpretujte míry přesnosti extrapolací "ex post" s horizontem jednoho období dopředu. Ověřte nekorelovanost chyb předpovědí "ex post" s horizontem jednoho období dopředu pomocí autokorelační funkce. Určete extrapolace na roky 2000 až 2002.



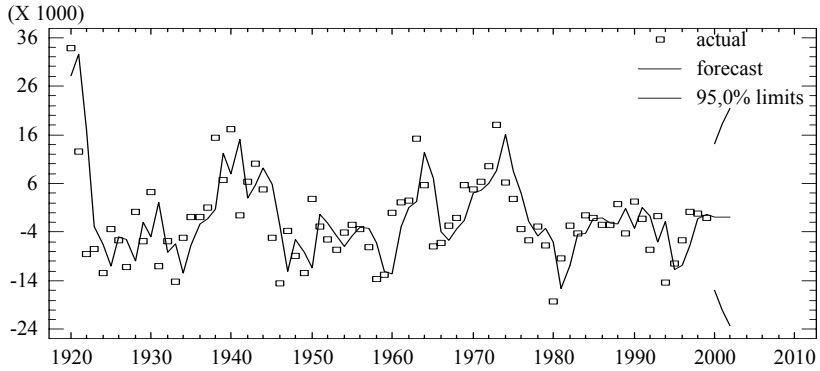
Obr. 2.11: Řada přírůstků (úbytků) počtu živě narozených dětí v ČR

Z obr. 2.11 je zřejmé, že řada přírůstků (úbytků) v období let 1921 až 1999 má stacionární charakter, takže můžeme zvolit model (2.60) ve tvaru

$$\text{diff(ZivNar_CR)}_t = \beta_0 + a_t.$$

Ve STATGRAPHICSu necháme volbou Optimize v doplňkovém panelu ***Model Specification Options*** najít nejvhodnější vyrovnávací konstantu pro Brownovo jednoduché exponenciální vyrovnávání.

Výsledek vyrovnávání nejvhodnější vyrovnávací konstantou $\alpha = 0{,}7799$ je na obr.2.11a.



Obr. 2.11a: Původní, vyrovnané a předpovězené hodnoty časové řady

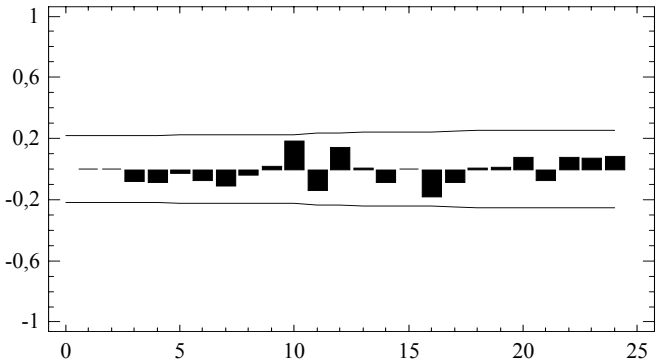
Průměrné charakteristiky chyb předpovědí "ex post" s horizontem $h = 1$ rok dopředu zjistíme z tab. 2.8a.

Tab. 2.8a. Průměrné charakteristiky chyb předpovědí ex post s h = 1

```
Data variable: diff(ZivNar_CR)                        Number of observations = 80
1 missing values were replaced with estimates
Start index = 1920
Sampling interval = 1,0 year(s)
Forecast model selected: Simple exponential smoothing with alpha = 0,7799
Number of forecasts generated: 3
Number of periods withheld for validation: 0

            Estimation      Validation
Statistic   Period          Period
-------------------------------------------
MSE         5,95932E7
MAE         5994,46
MAPE
ME          -465,03
MPE
```

Podle záporné hodnoty ME = –465,03 usuzujeme, že interpolace (v letech 1921 až 1999) modelem jednoduchého exponenciálního vyrovnávání systematicky nadhodnocuje skutečnost.

Volbou vybereme graf reziduální autokorelační funkce (Residual Autocorrelation Function), obr. 2.11b. Rezidua jsou v tomto případě chybami předpovědí "ex post" s horizontem jednoho roku dopředu a jsou určena vztahem $rezidua_t = diff(ZivNar_CR)_t - \hat{\beta}_{0,t-1}(1)$. Na základě grafu výběrové reziduální autokorelační funkce konstatujeme, že nesystematická složka nevykazuje autokorelaci a jednoduché exponenciální vyrovnávání přírůstků (úbytků) živě narozených dětí je tedy vyhovující.



Obr. 2. 11b: Reziduální ACF reziduí

Z vybereme Forecast Table a získáme tak předpovědi na jeden rok dopředu (tab. 2.8b).

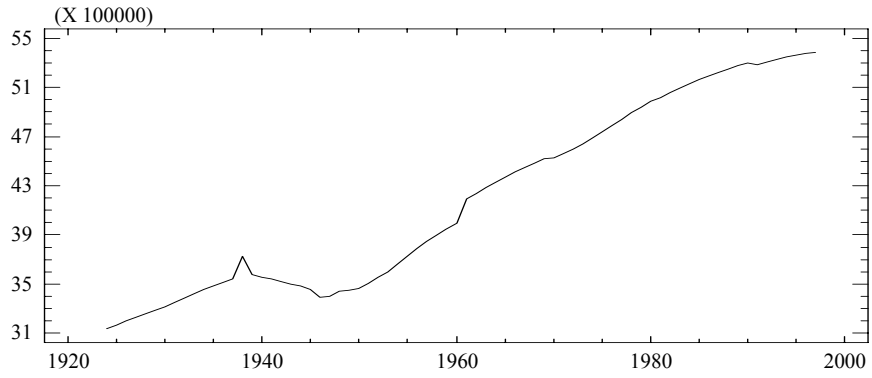
V první části zkrácené tabulky jsou uvedeny předpovědi "ex post" (tj. předpovědi určené v období analýzy časové řady). Ve druhé části jsou předpovědi "ex ante" (tj. předpovědi určené na počátku období předpovídání $T = 1999$ na roky 2000, 2001 a 2002 (pro horizonty $h = 1, 2, 3$).

Tab. 2.8b. Diferencovaná řada, předpovědi s $h = 1$ a 95% předpovědní interval

```
Forecast Table for diff(ZivNar_CR)
Model: Simple exponential smoothing with alpha = 0,7799
* = estimated
Period          Data                Forecast          Residual
-----------------------------------------------------------------
1920           *33779,0             28099,2             5679,8
1921            12613,0             32528,9           -19915,9
1922            -8553,0             16996,5           -25549,5
1923            -7498,0             -2929,5            -4568,4
1924           -12336,0             -6492,5            -5843,5
.
.
1985            -1060,0             -1337,9              277,9
1986            -2525,0             -1121,2            -1403,8
1987            -2435,0             -2216,0             -218,9
1988             1746,0             -2386,8             4132,8
1989            -4311,0               836,4            -5147,4
1990             2208,0             -3178,1             5386,1
1991            -1210,0              1022,5            -2232,5
1992            -7649,0              -718,6            -6930,4
1993             -680,0             -6123,6             5443,6
1994           -14446,0             -1878,1           -12567,9
1995           -10482,0            -11679,8             1197,8
1996            -5651,0            -10745,6             5094,6
1997              211,0             -6772,3             6983,3
1998             -122,0             -1326,0             1204,0
1999            -1064,0              -387,0             -676,9
-------------------------------------------------------------------
                                 Lower 95,0%        Upper 95,0%
Period          Forecast           Limit              Limit
-------------------------------------------------------------------
2000           -914,994           -15950,4            14120,4
2001           -914,994           -19982,4            18152,4
2002           -914,994           -23299,5            21469,6
```

Příklad 2.7

Časová řada středního stavu obyvatel Slovenska (Obyv_SR) v letech 1924 - 1997 je na obr. 2.12. Vyrovnejte časovou řadu Brownovým dvojitým exponenciálním vyrovnáváním a Holtovým exponenciálním vyrovnáváním a porovnejte jejich vlastnosti pomocí průměrných charakteristik chyb předpovědí. Zvolte posledních 6 let na verifikaci modelu. Verifikovaný model využijte na určení předpovědí s horizontem $h = 3$ roky, tj. na roky 1998 až 2000.



Obr. 2.12: Počet obyvatel Slovenska (střední stav)

Lze předpokládat, že časová řada počtu obyvatel Slovenska má vcelku lineární trend, jehož parametry se na lokálních úsecích mění v čase, a proto využijeme Brownův a Holtův model lineárního exponenciálního vyrovnávání.

Ve vstupním panelu **Forecasting** vyplníme název časové řady, počet předpovědí (Number of Forecasts) a délku období verifikace modelu (Withold for Validation). V doplňkovém panelu ***Model Specification Options*** zvolíme postupně nejprve Model A = Brown´s Linear Exp. Smoothing, potom Model B = Holt´s Linear Exp. Smoothing. V části **Parameters and Terms** zvolíme Optimize.

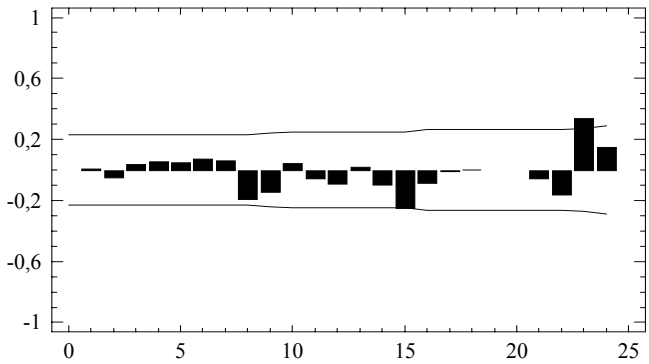
Z vybereme položku Model Comparisons, která umožňuje porovnávat současně několik modelů a rozhodnout se pro nejvhodnější z nich. Výsledky obou modelů jsou v tab. 2.9a.

Z tab. 2.9a vidíme, že průměrné charakteristiky chyb předpovědí jsou pro období interpolace (Estimation period) menší, použijeme-li Brownův model lineárního exponenciálního vyrovnávání s konstantou $\alpha = 0{,}5624$. Stejně tak v období verifikace modelu (Validation period) jsou průměrné charakteristiky při použití Brownova modelu lineárního exponenciálního vyrovnávání menší.

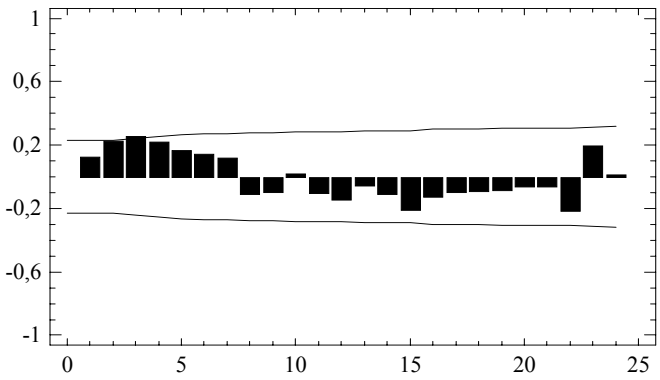
Tab. 2.9a. Průměrné charakteristiky chyb předpovědí

```
Model Comparison
Data variable: Obyv_SR                          Number of observations = 74
Start index = 1924                              Sampling interval = 1.0 year(s)
Number of periods withheld for validation: 6

Models
(A) Brown's linear exp. smoothing with alpha = 0,5624
(B) Holt's linear exp. smoothing with alpha = 0,9999 and beta = 0,0146

Estimation Period
Model    MSE            MAE            MAPE           ME             MPE           RMSE
-----------------------------------------------------------------------------------------
(A)   1,83385E9       23735,7       0,626157       2056,84        0,0813939      42823,5
(B)   1,9093E9        25324,0       0,657794       6880,0         0,152539       43695,5

Validation Period
Model    MSE            MAE            MAPE           ME             MPE
-----------------------------------------------------------------------------------------
(A)   3,70041E7      5235,63        0,0977601      172,407         0,00362091
(B)   1,83611E8      12516,9        0,233568       -12516,9       -0,233568
```

Analýzou reziduí Brownova modelu na obr. 2.12a a reziduí Holtova modelu na obr. 2.12b konstatujeme, že nesystematická složka není autokorelována. Statisticky významný koeficient autokorelace reziduí Brownova modelu ve zpoždění $k = 23$, lze považovat za nahodilý jev.



Obr. 2.12a: Reziduální ACF – Brownův model



Obr. 2.12b: ACF reziduí – Holtův model

Vzhledem k výše uvedeným skutečnostem použijeme k předpovídání Brownův model. Před výpočtem předpovědí je třeba ve STATGRAPHICSu ve vstupním panelu "vynulovat" políčko Withhold for Validation, abychom mohli počítat předpovědi na základě celé časové řady. Předpovědi počtu obyvatel Slovenska v letech 1998 až 2000 spolu s 95% intervaly spolehlivosti jsou v tab. 2.9b.

Tab. 2.9b: Předpovědi počtu obyvatel SR, Brownův model

```
Forecast Table for Obyv_SR
Model: Brown's linear exp. smoothing with alpha = 0,5618
V = withheld for validation 0
 ---------------------------------------------------------------------
                                   Lower 95,0%          Upper 95,0%
Period          Forecast              Limit                Limit
---------------------------------------------------------------------
1998            5,40613E6           5,3262E6            5,48607E6
1999            5,42156E6           5,30132E6           5,54179E6
2000            5,43698E6           5,27057E6           5,60339E6
```

2.2.7 Cvičení

1. Pomocí klouzavých průměrů různých délek vyrovnejte časovou řadu počtu obyvatel Slovenska (střední stav) v mil. osob (Obyv_SR) od roku 1924 do roku 1997 a pozorujte vliv délky klouzavé části na kvalitu vyrovnání časové řady.
2. Odhadněte pomocí klouzavých průměrů vhodné délky trend časové řady Obyv_SR v letech 1960 až 1997.
3. Pomocí koncových klouzavých průměrů určete předpovědi řady Obyv_SR na roky 1998, 1999 a 2000.
4. Určete pomocí Holtova exponenciálního vyrovnávání předpovědi řady Obyv_SR. Interpretujte charakteristiky MSE, MAE, MAPE, MPE a ME.
5. Vyrovnejte vhodnou trendovou funkcí časovou řadu Obyv_SR v letech 1960 až 1997. Odhadněte periodické-nesystematické výkyvy formou podílu skutečných hodnot a vyrovnaných hodnot.

2.3 Základní metody modelování sezónní složky, konstrukce předpovědí

V krátkodobých časových řadách měsíčních a čtvrtletních údajů se kromě trendové a nesystematické složky často vyskytuje také sezónní složka, obvykle s měsíční nebo čtvrtletní periodicitou. V časových řadách denních údajů se může vyskytovat různá periodicita, např. týdenní (každých sedm dní), dekádní (každých deset dní) apod.

Z grafického zobrazení časové řady není vždy jednoduché určit, zda řada obsahuje sezónnost a jaká je její periodicita. Pro ověření těchto skutečností se využívá periodogram nebo autokorelační funkce.

V následující části uvedeme způsob identifikace sezónnosti ve stacionárních časových řadách (řady bez trendu) a v nestacionárních časových řadách (v řadách s trendem).

2.3.1 Identifikace sezónnosti

Periodogram

Analýza časové řady pomocí periodogramu znamená rozklad časové řady na sinusové vlny (cykly - periody) s různými frekvencemi. Hodnoty periodogramu časové řady y_t pro $t = 1, 2, ..., T$ jsou definované vztahem

$$I(\omega_j) = \frac{1}{2}(a_j^2 + b_j^2),$$

kde

$$a_j = \frac{2}{T}\sum_{t=1}^{T} y_t \sin\omega_j t, \quad b_j = \frac{2}{T}\sum_{t=1}^{T} y_t \cos\omega_j t \quad \text{a} \quad \omega_j = \frac{2\pi j}{T}, \quad j = 1, 2, ..., T/2.$$

Frekvence ω se udává v radiánech za uvažovanou jednotku času, kterou je časový interval mezi dvěma sousedními pozorováními. Vysoké hodnoty periodogramu v jisté frekvenci ω_j indikují přítomnost periody (cyklu) určité délky.

Test sezónnosti pomocí autokorelační funkce

V části 2.2 jsme věnovali pozornost autokorelační funkci reziduí, přičemž jsme zdůraznili, že se využívá pro ověřování neautokorelovanosti nesystematické složky modelů trendů. Její praktické použití je však širší a jak dále zjistíme, dá se používat také pro ověřování sezónnosti v časových řadách. Nechť časová řada y_t pro $t = 1, 2, ..., T$ je stacionární a má sezónnost délky s sezón. Otázkou je, zda je sezónnost identifikovatelná.

Jsou-li časové řady dostatečně dlouhé, určujeme výběrovou autokorelační funkci na základě časově zpožděných řad y_t a y_{t-k} podle vztahu

$$r_k = \hat{\rho}_k = \frac{\sum_{t=k+1}^{T}(y_t - \bar{y})(y_{t-k} - \bar{y})}{\sum_{t=1}^{T}(y_t - \bar{y})^2} \in \langle -1, 1\rangle .$$

Na základě této funkce zjišťujeme charakter autokorelace stochastického procesu. Testovanou a alternativní hypotézu formulujme takto

$$H_0: \rho_k = 0$$

$$H_1: \rho_k \neq 0.$$

Testové kritérium $U = r_k \sqrt{T}$ má podle empirického pravidla přibližně normované normální rozdělení, takže na 5% hladině významnosti zamítáme nulovou hypotézu platí-li

$$|r_k| > u_{0,025} \frac{1}{\sqrt{T}} = \frac{1{,}96}{\sqrt{T}},$$

tj. se spolehlivostí 95 % tvrdíme, že v časové řadě existuje autokorelace se zpožděním k období.

Empirické pravidlo je vhodným testem sezónnosti určitého typu pro délku s sezón za předpokladu, že testujeme koeficient autokorelace veličin y_t a y_{t-s}. Je-li uvedený korelační koeficient r_k na 5% hladině významnosti statisticky významně odlišný od nuly tvrdíme, že v řadě existuje statisticky významná sezónnost s počtem s sezón.

Poznámka:

Zjišťujeme-li pomocí autokorelační funkce statistickou významnost sezónnosti v časových řadách s trendem, musíme nejprve tyto řady stacionarizovat (viz Boxova-Jenkinsova metodologie).

2.3.2 Identifikace sezónnosti ve STATGRAPHICSu

Identifikaci sezónní složky pomocí periodogramu a autokorelační funkce provádíme ve STATGRAPHICSu v proceduře **Descriptive Methods**

Special
⇒ **Time-Series Analysis**
⇒ **Descriptive Methods**.

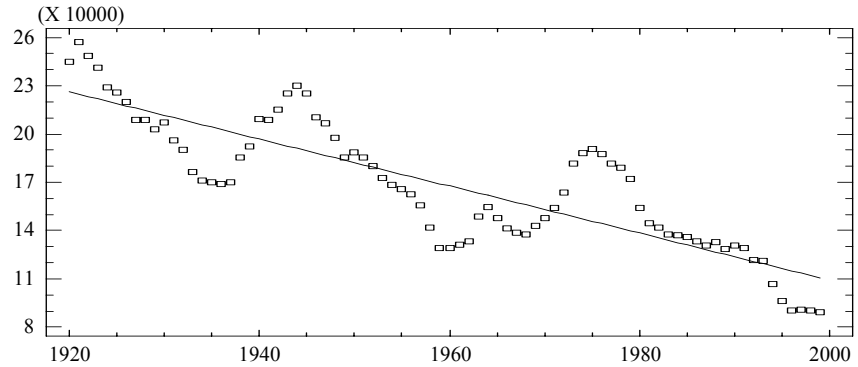
Po vyplnění vstupního panelu nejprve dáváme přednost analýze grafů, takže z nabídky vybereme položky Horizontal Time Sequence Plot, Periodogram a Autocorrelation Function. Potřebujeme-li znát i hodnoty periodogramu nebo hodnoty koeficientů autokorelací, vybíráme z nabídky položky Autocorrelations a Periodogram Table.

Příklad 2.8

Z příkladu 2.4 víme, že vývoj počtu živě narozených dětí v České republice (ZivNar_CR) v letech 1920 - 1999 (obr. 2.9) má klesající trend a výraznou cyklickou složku. Předpokládáme tedy model $y_t = T_t + C_t + a_t$. Odhadněte trend řady lineární funkcí a určete řadu očištěnou od lineárního trendu, která bude obsahovat cyklicko-nesystematickou složku. Periodogramem najděte cykly a zjistěte jejich délku.

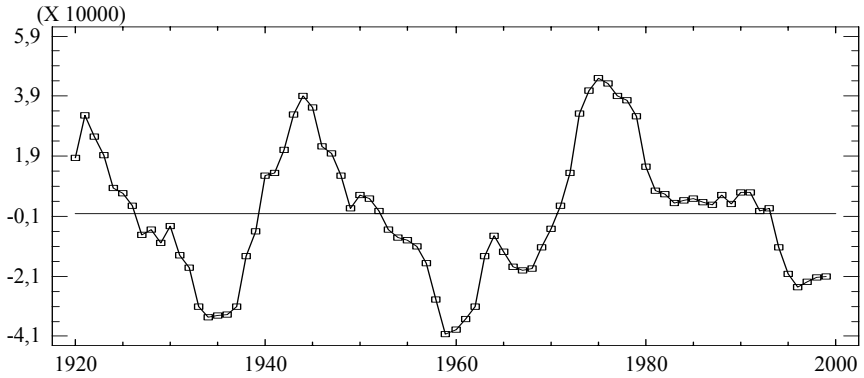
Odhad lineárního trendu počtu živě narozených dětí je na obr. 2.13 a má tvar:

$$ZivNar_CR_t = 227684{,}0 - 1463{,}69t \qquad \text{pro } t = 1, 2, ..., 80.$$



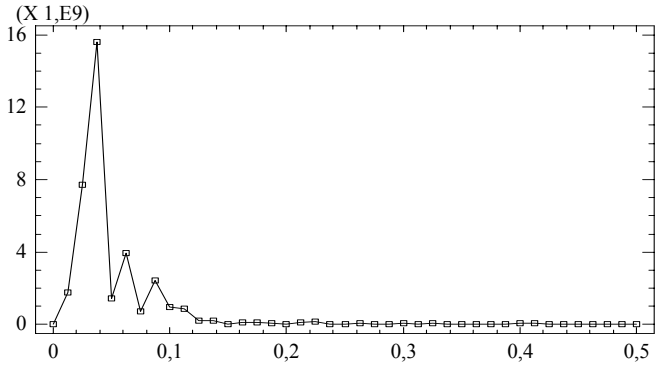
Obr. 2.13: Lineární trend počtu živě narozených dětí v ČR v letech 1920 - 1999

Graf reziduí, tj. řada cyklicko-nesystematických výkyvů je na obr. 2.13a.



Obr. 2.13a: Cyklicko-nesystematické výkyvy kolem lineárního trendu počtu živě narozených dětí

Periodogram časové řady očištěné od lineárního trendu je na obr. 2.13b.



Obr. 2.13b: Periodogram řady reziduí (řada očištěná od lineárního trendu)

V periodogramu vidíme tři vrcholy, je proto možné očekávat tři cykly s různou délkou. Z grafu se frekvence jednotlivých vrcholů nedají přesně zjistit, proto použijeme hodnoty periodogramu (tab. 2.10).

Maximální hodnota je ve frekvenci 0,0375, což signalizuje přítomnost periody délky asi 27 let (viz sloupec `Period`). V řadě se dále vyskytuje druhý vrchol s frekvencí 0,0625, což odpovídá délce cyklu 16 let, frekvence třetího vrcholu je 0,0875 odpovídající délce cyklu 11 až 12 let.

Tab. 2.10. Hodnoty periodogramu reziduí lineárního trendu živě narozených dětí

Periodogram for residuals
Data variable: ZivNar_CR
Model: Linear trend = 3,0365E6 -1463,69 t

Frequency	Period	Ordinate	Cumulative Sum	Integrated Periodogram
0,0		1,38078E-19	1,38078E-19	3,73233E-30
0,0125	80,0	1,76194E9	1,76194E9	0,0476262
0,025	40,0	7,73218E9	9,49412E9	0,256631
0,0375	**26,67**	**1,56126E10**	2,51067E10	0,678648
0,05	20,0	1,43287E9	2,65396E10	0,71738
0,0625	**16,0**	**3,92285E9**	3,04624E10	0,823416
0,075	13,333	7,50556E8	3,1213E10	0,843704
0,0875	**11,4286**	**2,42309E9**	3,36361E10	0,909202
0,1	10,0	9,50419E8	3,45865E10	0,934892
0,1125	8,88889	8,56656E8	3,54432E10	0,958048
0,125	8,0	2,04994E8	3,56482E10	0,963589
0,1375	7,27273	2,20691E8	3,58689E10	0,969554
0,15	6,66667	2,39833E7	3,58928E10	0,970203
0,1625	6,15385	1,35965E8	3,60288E10	0,973878
0,175	5,71429	1,0837E8	3,61372E10	0,976807
0,1875	5,33333	7,18676E7	3,6209E10	0,97875
0,2	5,0	2,2112E7	3,62311E10	0,979347
0,2125	4,70588	1,14034E8	3,63452E10	0,98243
0,225	4,44444	1,59387E8	3,65046E10	0,986738
0,2375	4,21053	1,92755E7	3,65238E10	0,987259
0,25	4,0	2,01753E7	3,6544E10	0,987805
0,2625	3,80952	7,35926E7	3,66176E10	0,989794
0,275	3,63636	3,04434E7	3,66481E10	0,990617
0,2875	3,47826	4,59577E6	3,66527E10	0,990741
0,3	3,33333	5,12861E7	3,67039E10	0,992127
0,3125	3,2	5,76715E6	3,67097E10	0,992283
0,325	3,07692	6,5444E7	3,67751E10	0,994052
0,3375	2,96296	1,94333E7	3,67946E10	0,994577
0,35	2,85714	2,69775E6	3,67973E10	0,99465
0,3625	2,75862	4,25049E6	3,68015E10	0,994765
0,375	2,66667	1,36008E7	3,68151E10	0,995133
0,3875	2,58065	7,37326E6	3,68225E10	0,995332
0,4	2,5	6,66383E7	3,68891E10	0,997133
0,4125	2,42424	4,74743E7	3,69366E10	0,998417
0,425	2,35294	2,60738E6	3,69392E10	0,998487
0,4375	2,28571	1,01066E7	3,69493E10	0,99876
0,45	2,22222	1,18961E6	3,69505E10	0,998792
0,4625	2,16216	4,6399E6	3,69552E10	0,998918
0,475	2,10526	1,57162E6	3,69567E10	0,99896
0,4875	2,05128	3,01075E7	3,69868E10	0,999774
0,5	2,0	8,35437E6	3,69952E10	1,0

2.3.3 Sezónní dekompozice, určení sezónních indexů

Je-li v časové řadě identifikována sezónnost, je třeba ji modelovat. Nejstarší metodou odhadu sezónních výkyvů je metoda dekompozice. Odhad jednotlivých složek se provádí postupně, pro každou složku zvlášť. Trend se odhaduje jednoduchými klouzavými průměry, je-li délka sezónnosti liché číslo, např. $s = 7$ dní, nebo váženými (centrovanými) klouzavými průměry, je-li délka sezónnosti sudé číslo, např. $s = 12$ nebo 4. Sezónní složka se odhaduje pomocí sezónních průměrů (aditivní model) nebo pomocí sezónních indexů (multiplikativní model). Sezónní průměry interpretujeme ve stejných měrných jednotkách

jako původní časovou řadu, sezónní indexy se interpretují v procentech. Sezónní dekompozice se využívá především na získání odhadů sezónních výkyvů a sezónně očištěné časové řady.

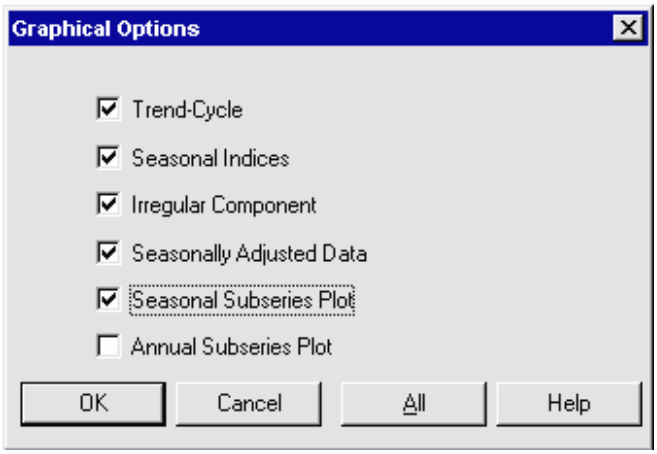
2.3.4 Sezónní dekompozice ve STATGRAPHICSu

Metodu sezónní dekompozice najdeme v proceduře **Seasonal Decomposition**

Special
⇒ **Time-Series Analysis**
⇒ **Seasonal Decomposition**.

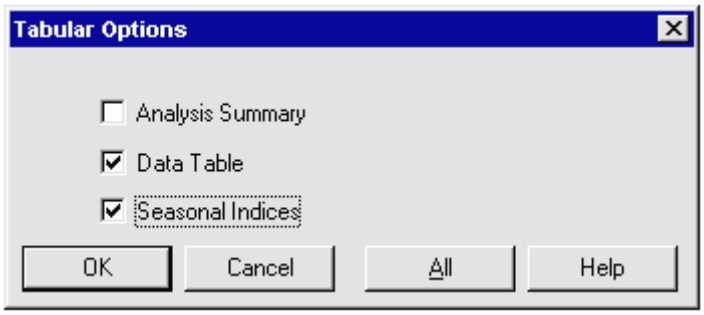
Ve vstupním panelu ***Seasonal Decomposition*** specifikujeme základní vlastnosti časové řady.

K identifikaci jednotlivých složek získaných ze sezónní dekompozice slouží nabídky výpočetních metod (▤) a grafů (▣).



- Trend-cyklická složka (klouz. prům.)
- Sezónní indexy (průměry)
- Reziduální složka
- Sezónně očištěná časová řada
- Graf řady uspořádaný podle sezón
- Graf řady uspořádaný podle let

Obr. 2.14a: Nabídka grafů sezónní dekompozice

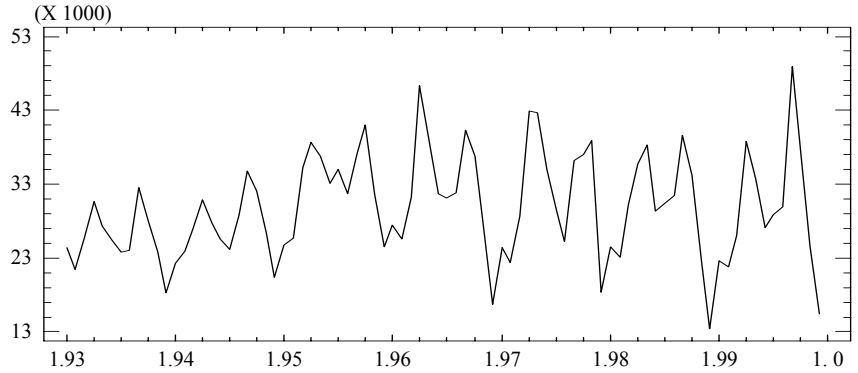


- Informace o časové řadě
- Rozklad časové řady na složky
- Sezónní indexy

Obr. 2.14b: Nabídka výpočetních metod sezónní dekompozice

Příklad 2.9

Je dána měsíční časová řada počtu odregistrovaných uchazečů o zaměstnání na Slovensku (NezOdr_SR) od ledna 1993 do prosince 1999 v tis. osob. Vývoj tohoto ukazatele je na obr. 2.15. Sezónní dekompozicí odhadněte jednotlivé složky časové řady a určete sezónně očištěnou časovou řadu.



Obr. 2.15: Počet odregistrovaných uchazečů o zaměstnání na Slovensku 1/1993-12/2000

Z obr. 2.15 vidíme, že variabilita hodnot v časové řadě v čase roste, proto použijeme multiplikativní model $y_t = TC_t \,.\, S_t \,.\, I_t$. Trend je díky sezónnosti nejasný, může být nejprve mírně rostoucí (do roku 1996) a poté spíše klesající.

Výsledky multiplikativní sezónní dekompozice jsou v tab. 2.11a a vyžadují komentář. Ve sloupci `Trend-Cycle` jsou centrované klouzavé průměry, které jsou určeny podle vztahu

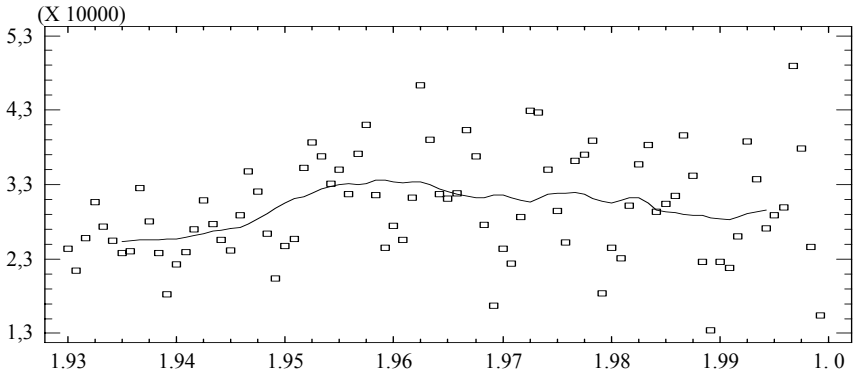
$$CKP_t = \frac{1}{24}(y_{t-6} + 2y_{t-5} + 2y_{t-4} + 2y_{t-3} + 2y_{t-2} + 2y_{t-1} + 2y_t + \\ + 2y_{t+1} + 2y_{t+2} + 2y_{t+3} + 2y_{t+4} + 2y_{t+5} + y_{t+6}) \qquad t = 7, 8, ..., T-6.$$

Tab. 2.11a. Výsledky multiplikativní sezónní dekompozice

Seasonal decomposition method: Multiplicative

Period	Data	Trend-Cycle	Seasonality	Irregular	Seasonally Adjusted
1/93	24370.0				29914.8
2/93	21476.0				27027.0
3/93	25805.0				26143.5
4/93	30676.0				23979.1
5/93	27380.0				22974.4
6/93	25444.0				25575.8
7/93	23849.0	25379.2	93.9706	96.4603	24480.9
8/93	24035.0	25394.0	94.6482	97.0021	24632.8
9/93	32527.0	25545.5	127.329	102.37	26151.0
10/93	28001.0	25602.3	109.369	93.5081	23940.2
11/93	23781.0	25619.2	92.8249	97.5264	24985.5
12/93	18244.0	25634.9	71.1687	114.302	29301.1
-					
-					
1/99	22662.0	28407.8	79.774	97.9245	27818.1
2/99	21769.0	28280.5	76.9752	96.8712	27395.7
3/99	25981.0	28604.0	90.8301	92.0216	26321.8
4/99	38804.0	29142.6	133.152	104.083	30332.6
5/99	33685.0	29379.6	114.654	96.2056	28264.8
6/99	27140.0	29547.4	91.8525	92.3283	27280.6
7/99	28885.0				29650.3
8/99	29936.0				30680.5
9/99	48889.0				39305.7
10/99	37795.0				32313.8
11/99	24606.0				25852.3
12/99	15405.0				24741.5

Centrované klouzavé průměry vyjadřují odhad trendové-cyklické složky a její grafické vyjádření je na obr. 2.15a.



Obr. 2.15a: Trendová-cyklická složka

V tab. 2.11a, ve sloupci `Seasonality`, jsou uvedeny odhady sezónní–nesystematické složky v procentech, které získáme jako podíl hodnot časové řady a centrovaných klouzavých průměrů, tj.

$$\hat{S_t I_t} = \frac{y_t}{CKP_t} \,.\, 100\ \%, \qquad t = 7, 8, ..., T - 6.$$

Např. pro $t = 10$, tj. pro říjen roku 1993

(28001/25602,3).100 % = 109,369 %.

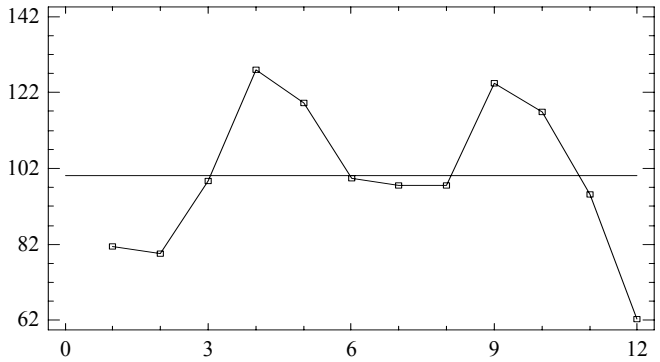
Poznámka:

Odhad sezónních výkyvů, tj. sezónní indexy, určujeme dodatečně a to průměrováním hodnot ze sloupce `Seasonality` za stejné sezóny.

Průměrné sezónní indexy z výstupu STATGRAPHICSu jsou v tab. 2.11b a jejich graf na obr. 2.15b.

Tab. 2.11b. Sezónní indexy ($\hat{S}_t$)

```
Seasonal Indices for NezOdr_SR
Seasonal decomposition method: Multiplicative
Season       Index
------------------------
1            81.4648
2            79.4614
3            98.7052
4           127.9280
5           119.1760
6            99.4846
7            97.4190
8            97.5734
9           124.3820
10          116.9620
11           95.1792
12           62.2639
```



Obr. 2.15b: Sezónní indexy ($\hat{S}_t$)

Z tab. 2.11b a obr.2.15b vidíme, že každoročně nejvyšší počet odregistrovaných je v dubnu (začátek sezónních prací) a v září (nástup studentů do škol). Nejmenší počet odregistrovaných je v prosinci (důsledek zimního období a útlumu sezónních prací).

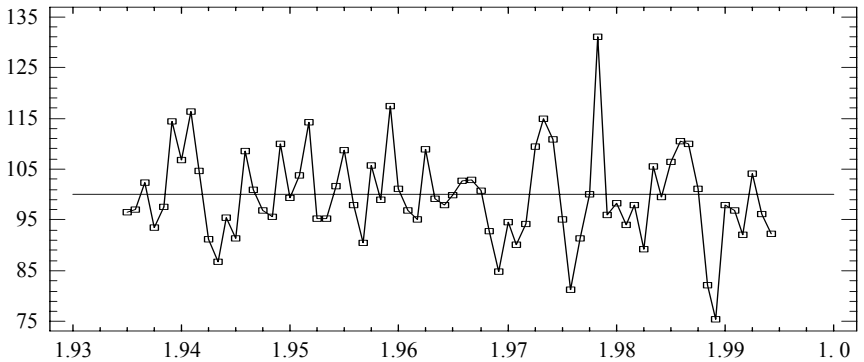
V tab. 2.11a, ve sloupci `Irregular`, jsou uvedena rezidua, tj. odhady nesystematické složky v procentech, které získáme jako podíl hodnot časové řady a součinu centrovaných klouzavých průměrů a sezónních indexů, tj.

$$\hat{I}_t = \frac{y_t}{CKP_t \cdot \hat{S}_t} \,.\, 100\ \%, \qquad t = 7, 8, ..., T - 6.$$

Např. pro $t = 10$, tj. pro říjen roku 1993 platí

(28001/(25602,3*1,16962)).100 % = (28001/29934,201).100 % = 93,5081 %.

Tato rezidua jsou zachycena na obr. 2.15c.

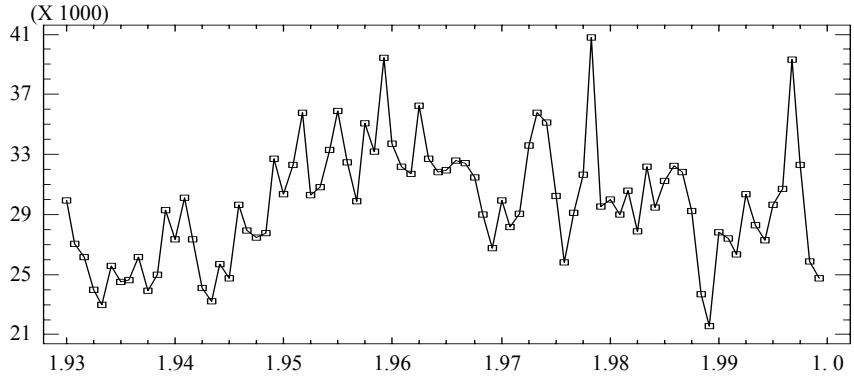


Obr. 2.15c: Rezidua ($\hat{I}_t$)

Nakonec v tab. 2.11a, ve sloupci `Seasonally Adjusted` se nacházejí hodnoty sezónně očištěné časové řady, které získáme jako podíly hodnot časové řady a příslušných sezónních indexů

$$SAy_t = \frac{y_t}{\hat{S}_t} \,.$$

Tato sezónně očištěná časová řada je zachycena na obr. 2.15d a lze ji chápat jako odhad trendové-nesystematické složky.



Obr. 2.15d: Sezónně očištěná časová řada

2.3.5 Předpovídání časových řad po sezónní dekompozici

Odhad trendu centrovanými klouzavými průměry považujeme za předběžný odhad, který není vhodný pro předpovídání. Řada klouzavých průměrů je kratší o $s/2$ prvních a $s/2$ posledních hodnot, a bylo by třeba ještě předpovídat poslední hodnoty časové řady, které jsou ve skutečnosti známé.

Trendovou složku původní řady proto předpovídáme ze sezónně očištěné řady. Na jejím základě zvolíme vhodný model trendu (vhodnou funkci času nebo model exponenciálního vyrovnávání) a odhadneme jeho parametry. Jeho prostřednictvím potom určíme předpovědi trendu v horizontech $h = 1, 2, \ldots$.

Předpovědi původní časové řady se nakonec vypočítají tak, že se předpovědi trendu vynásobí sezónními indexy (vydělenými stem), nebo se k předpovědím trendu přičtou sezónní průměry, podle toho, jaký model jsme v sezónní dekompozici zvolili.

Příklad 2.10

Určete předpovědi počtu odregistrovaných uchazečů o zaměstnání na Slovensku na měsíce leden až březen roku 2000 za předpokladu, že využijete výsledky sezónní dekompozice z příkladu 2.9. Hodnoty sezónně očištěné časové řady jsou v příloze (SA_NezOdr_SR).

Z obr. 2.15d je zřejmé, že jako model trendu této sezónně očištěné časové řady lze zvolit nelineární trendovou funkci (polynom vyššího řádu) nebo exponenciální vyrovnávání. Pokud si uvědomíme, že časovou řadu můžeme rozdělit na lokální úseky s konstantní úrovní, jejíž hodnoty se mění v čase, nebo na lokální úseky s lineárním trendem, potom vhodné modely vyrovnávání a předpovídání mohou být jednoduché Brownovo exponenciální vyrovnávání nebo Holtův model lineárního trendu.

V proceduře **Forecasting** v doplňkovém panelu ***Model Specification Options*** zvolíme jako model A jednoduché exponenciální vyrovnávání (Simple Exp. Smoothing), které aplikujeme na sezónně očištěnou časovou řadu. Stejně určíme výsledky podle modelu B, kde uplatníme Holtův model exponenciálního vyrovnávání (Holt´s Linear Exp. Smoothing). Dále vybereme z Model Comparisons a porovnáme kvalitu výsledků vyrovnávání oběma modely na základě průměrných měr přesnosti.

Tab. 2.12a. Výsledky exponenciálního vyrovnávání sezónně očištěné řady

```
Model Comparison
Data variable: SA_NezOdr_SR          Number of observations = 84
Start index = 1/93                   Sampling interval = 1.0 month(s)

Models
(A) Simple exponential smoothing with alpha = 0.3431
(B) Holt's linear exp. smoothing with alpha = 0.3521 and beta = 0.0236

Estimation Period
Model     MSE          MAE        MAPE         ME           MPE
--------------------------------------------------------------------------
(A)    1,07965E7    2455,45      8,25149      42,9360      -0,811402
(B)    1,12014E7    2511,23      8,42493      90,8947      -0,596428
```

Podle tab. 2.12a jsou průměrné charakteristiky chyb předpovědí pro $h = 1$ měsíc dopředu o něco příznivější pro jednoduché exponenciální vyrovnávání, než pro Holtovo lineární exponenciální vyrovnávání. Jako prognostický model proto volíme jednoduché exponenciální vyrovnávání. Podle míry MAPE lze očekávat, že extrapolace určíme asi s 8,25% chybou. Kromě toho, z kladné a vysoké hodnoty ME = 42,936 usuzujeme, že vyrovnané hodnoty trendu modelem A budou podhodnocené, a tedy i výsledné předpovědi sezónně očištěné řady budou pravděpodobně podhodnocené.

Výsledky předpovědí sezónně očištěné časové řady modelem A spolu se sezónními indexy $\hat{S}_j$, předpověďmi, skutečnými hodnotami a chybami předpovědí na leden až březen roku 2000 jsou v tab. 2.12b. Předpovědi původní časové řady určíme jako součin předpovědí sezónně očištěné řady a sezónních indexů.

Tab. 2.12b. Předpovědi časové řady - Simple Exp. Smoothing

Měsíc	Předpovědi sez. očišť. č. ř.	Sezónní indexy	Předpověď pův. č. ř.	Původní č. ř.	Chyba
Leden	28359,4	0,81464	23102,9	27944	4841,1
Únor	28359,4	0,79461	22534,8	26952	4417,2
Březen	28359,4	0,98705	27992,2	32354	4361,8

Podle posledního sloupce tab. 2.12b vidíme, že předpovědi jsou skutečně podhodnocené, přibližně s MAPE = 15% chybou.

2.3.6 Regresní metoda modelování sezónnosti

V části 2.3.4 jsme pro předpovídání sezónní časové řady použili kombinaci dvou technik: sezónní dekompozice a modelů trendu. Předpověď původní řady jsme potom vypočítali jako součin (nebo součet) předpovědí trendu sezónně očištěné řady a sezónního výkyvu.

Pro konstrukci předpovědí sezónní časové řady je možné využít také regresní model s umělými proměnnými, ve kterém odhadneme parametry trendu a sezónnosti současně. Předpokládejme aditivní model časové řady

$$y_t = T_t + S_t + a_t,$$

ve kterém a_t jsou náhodné veličiny tvořící řadu typu bílého šumu.

Trendovou složku T_t modelujeme vhodnou funkcí času, např. přímkou, parabolou, hyperbolou. Sezónní složku S_t vyjadřujeme pomocí umělých (nula – jedničkových) proměnných ("dummy variables"), které přiřazují hodnotě časové řady jedničku, nachází-li se v uvažované sezóně a nulu jinak.

Jde-li o dvanáctiměsíční sezónnost, lze použít dvanáct umělých proměnných D_j, pro $j = 1, 2, ..., 12$, které mají následující hodnoty:

$D_1 = (1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0) \Rightarrow$ pro lednové hodnoty řady,

$D_2 = (0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0) \Rightarrow$ pro únorové hodnoty řady,

$D_3 = (0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0) \Rightarrow$ pro březnové hodnoty řady,

...

$D_{12} = (0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1) \Rightarrow$ pro prosincové hodnoty řady.

V případě čtvrtletní sezónnosti lze použít čtyři proměnné, D_j $j = 1, 2, ..., 4$, které mají následující hodnoty:

$D_1 = (1, 0, 0, 0)$, $D_2 = (0, 1, 0, 0)$, $D_3 = (0, 0, 1, 0)$, $D_4 = (0, 0, 0, 1)$.

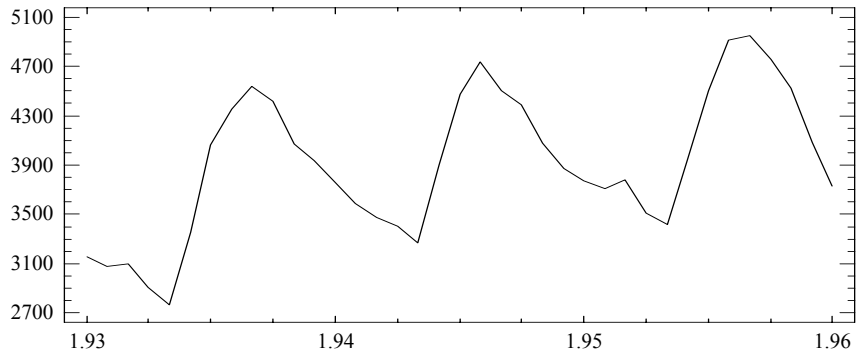
Poznámka:

1. V případě přítomnosti volného parametru (konstanty) v modelu trendu je počet umělých proměnných v regresním modelu se sezónností délky s sezón maximálně $s - 1$, aby se zabránilo multikolinearitě.
2. Ověřování vhodnosti regresního modelu je analogické jako v kterémkoli regresním modelu. Důležité je zejména testování autokorelace a heteroskedasticity nesystematické složky.
3. Předpovídání časové řady pomocí regresního modelu vyžaduje zvolit hodnoty časové proměnné v horizontech $h > 0$ a pro sezónní proměnné dosadit jedničky příslušných sezón v horizontu h.

Příklad 2.11

Jsou dány měsíční hodnoty počtu registrovaných nezaměstnaných absolventů gymnázií na Slovensku (NezGym_SR) od ledna 1993 do prosince 1996. Graficky analyzujte vývoj této časové řady v letech 1993 až 1995. Zvolte vhodný regresní model a otestujte jeho kvalitu. Určete předpovědi na rok 1996 a porovnejte je se skutečností.

Časovou řadu počtu registrovaných nezaměstnaných absolventů gymnázií na Slovensku od ledna 1993 do prosince 1995 obsahuje obr. 2.16.



Obr. 2.16: Vývoj počtu nezaměstnaných absolventův gymnázií v SR v letech 1993-1995

Z grafu je zřejmé, že časová řada má v uvedeném úseku rostoucí, přibližně lineární trend a dvanáctiměsíční sezónnost. Navrhovaný regresní model řady má tvar:

$$y_t = \beta_0 + \beta_1 t + \beta_2 D_2 + \beta_3 D_3 + \beta_4 D_4 + \ldots + \beta_{12} D_{12} + a_t.$$

Odhad parametrů modelu je v tab. 2.13a).

Tab. 2.13a) Odhad regresního modelu pro zkrácenou řadu NezGym_SR

Multiple Regression Analysis

Dependent variable: take(NezGym_SR,36)

Parameter	Estimate	Standard Error	T Statistic	P-Value
CONSTANT	3284,26	86,1915	38,1043	0,0000
t	21,2361	2,37144	8,95493	0,0000
D2	-122,903	113,854	-1,07948	0,2916
D3	-153,472	113,928	-1,3471	0,1911
D4	-351,708	114,051	-3,08377	0,0052
D5	-492,944	114,224	-4,3156	0,0003
D6	66,1528	114,445	0,57803	0,5689
D7	660,25	114,715	5,75556	0,0000
D8	956,347	115,033	8,31365	0,0000
D9	932,444	115,399	8,08014	0,0000
D10	769,208	115,813	6,64181	0,0000
D11	451,639	116,273	3,88428	0,0007
D12	169,403	116,78	1,45061	0,1604

t-test indikuje, že parametry lineárního trendu a sezónní výkyvy v dubnu, květnu, červenci, srpnu, září, říjnu a listopadu jsou statisticky významné na 5% hladině významnosti. Naopak sezónní výkyvy v únoru, březnu, červnu a prosinci nejsou na 5% hladině významnosti statisticky významné.

Poznámka:

Nevýhodou regresního modelu s umělými proměnnými je, že přímo nepoznáme konkrétní hodnoty sezónních výkyvů (sezónní průměry pro jednotlivé měsíce). Lze je dopočítat následujícím způsobem.

1. Určíme součet odhadů všech sezónních regresních parametrů a vydělíme je počtem sezón. Je-li počet sezón 12

$$\hat{s} = \frac{\hat{\beta}_2 + \hat{\beta}_3 + \hat{\beta}_4 + \ldots + \hat{\beta}_{12}}{12}.$$

   Záporná hodnota průměru sezónních regresních parametrů je odhadem průměrného sezónního výkyvu v lednu, tj. $-\hat{s} = \hat{s}_1$.

   V našem příkladu je $\hat{s} = 240{,}368$, což interpretujeme tak, že se každoročně v lednu průměrný počet registrovaných absolventů gymnázií snižoval oproti trendu o 241 osob.

2. Průměrné sezónní výkyvy ostatních sezón $j = 2, 3, \ldots, 12$ určíme vztahem

$$s_j = \hat{\beta}_j - \hat{s},$$

   tj.

   $\hat{s}_2 = -363{,}271$, $\hat{s}_3 = -393{,}840$, $\hat{s}_4 = -592{,}076$, $\hat{s}_5 = -733{,}312$,

   $\hat{s}_6 = -174{,}215$, $\hat{s}_7 = 419{,}882$, $\hat{s}_8 = 715{,}979$, $\hat{s}_9 = 692{,}076$,

   $\hat{s}_{10} = 528{,}840$, $\hat{s}_{11} = 211{,}271$, $\hat{s}_{12} = -70{,}965$.

3. V tomto případě musíme upravit také konstantní člen modelu a to tak, že k jeho původnímu odhadu připočítáme průměr sezónních regresních koeficientů. Odhad trendu potom má tvar

$$\hat{T}_t = (\hat{\beta}_0 + \hat{s}) + \hat{\beta}_1 t = 3524{,}628 + 21{,}2361t.$$

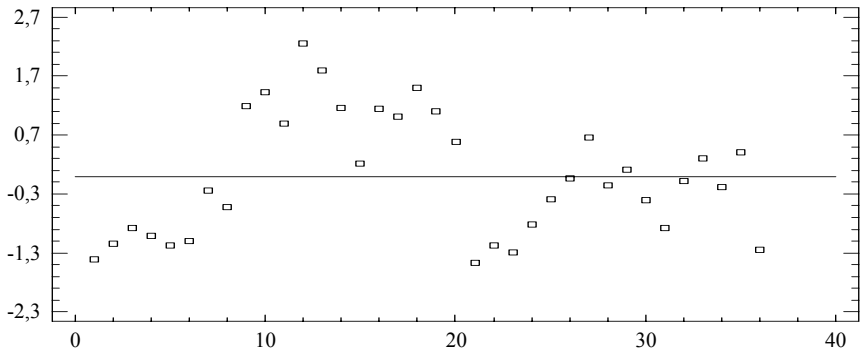
Stejným způsobem postupujeme, pro čtvrtletní údaje.

Pro posouzení kvality odhadu regresního modelu využijeme výstup analýzy rozptylu v tab. 2.13b).

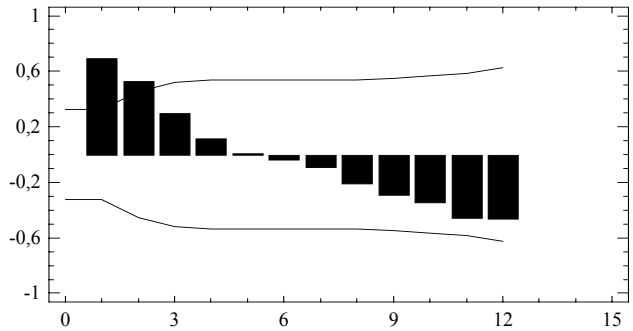
Tab. 2.13b. Analýza rozptylu odhadu lineárního trendu s umělými proměnnými

```
                        Analysis of Variance
--------------------------------------------------------------------------------
Source        Sum of Squares      Df      Mean Square    F-Ratio      P-Value
--------------------------------------------------------------------------------
Model           1,16178E7         12      968149,0         49,81        0,0000
Residual        447021,00         23      19435,7
--------------------------------------------------------------------------------
Total (Corr.)      1.20648E7      35

R-squared = 96,2948 percent
R-squared (adjusted for d.f.) = 94,3617 percent
Standard Error of Est. = 139,412
Mean absolute error = 96,0741
Durbin-Watson statistic = 0,523636
```

Z hodnoty indexu determinace `R-squared = 96,2948 %` usuzujeme, že model vysvětlil variabilitu počtu refistrovaných nezaměstnaných z 96,2948 %. Protože Durbinova-Watsonova statistika DW = 0,523636 < 2 tvrdíme, že DW test prokázal na 5% hladině významnosti hypotézu, že nesystematická složka modelu je autokorelovaná. O autokorelaci nesystematické složky se přesvědčíme také graficky z obr. 2.16a a z autokorelační funkce reziduí na obr. 2.16b.



Obr. 2.16a: Rezidua regresního modelu



Obr. 2.16b: ACF reziduí regresního modelu

Z obr. 2.16b je zřejmé, že autokorelační koeficient ve zpoždění 1 měsíc je statisticky významný na 5% hladině významnosti. Z toho vyplývá, že odhadnutý model není kvalitní a musíme jej změnit.

Regresní modely odhadujeme ve STATGRAPHICSu v proceduře

Relate
⇒ **Multiple Regression**.

Vstupní panel vyplníme při použití datového souboru *Data.sf3* následujícím způsobem. Časová řada NezGym_SR má v období od ledna 1993 do prosince 1996 48 hodnot. Abychom brali v úvahu jen hodnoty v letech 1993 – 1995 zkrátíme původní časovou řadu pomocí příslušného operátoru tak, že do prvního řádku vstupního panelu vícenásobné regresní analýzy (Dependent Variable:) vložíme

TAKE(NezGym_SR;36)

což znamená, že se z původní časové řady vybere pouze prvních 36 hodnot.

Dále připravíme časovou proměnnou t a jednotlivé sezónní umělé proměnné D_2 až D_{12} (proměnnou D_1 vynecháme), které obsahují 48 hodnot (stejně jako původní časová řada), aby je bylo možné použít pro konstrukci předpovědí na rok 1996. Hodnoty těchto proměnných lze vypsat buď z klávesnice do tabulkového editoru STATGRAPHICSu nebo vygenerovat pomocí operátorů v panelu ***Generate Data***, kde do příkazového řádku Expression: zadáváme

pro časovou proměnnou, obsahující hodnoty od 1 do 48

t COUNT(1;48;1)

a pro každou z 11 sezónních umělých proměnných, obsahující v j-té sezóně 1, v ostatních 0

D_2 RESHAPE(TAKE(ROWS(2;2);12);48)

D_3 RESHAPE(TAKE(ROWS(3;3);12);48)

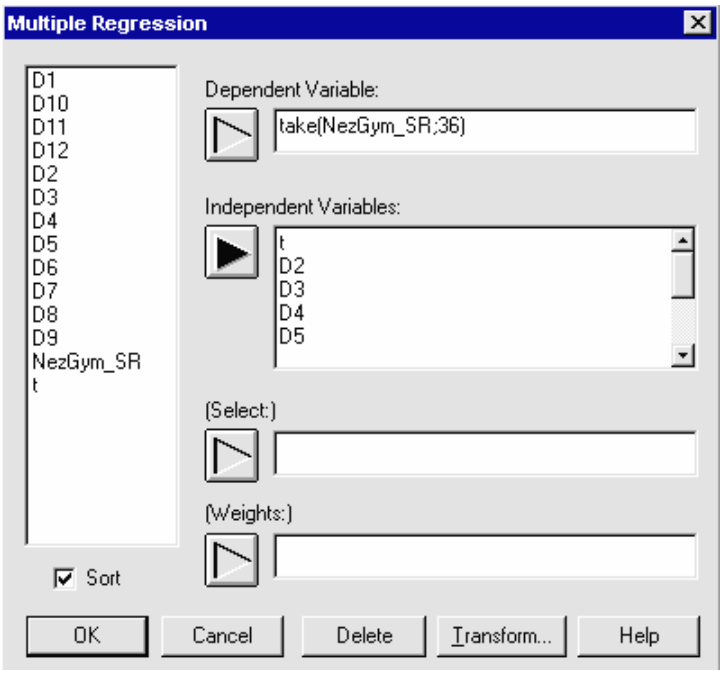
D_4 RESHAPE(TAKE(ROWS(4;4);12);48)

...

D_{12} RESHAPE(TAKE(ROWS(12;12);12);48)

(seznam použitých operátorů viz Příloha).

Všechny takto připravené proměnné vložíme do oblasti Independent Variables: ve vstupním panelu pod sebe (obr. 2.16c).



Obr. 2.16c: Vyplnění vstupního panelu vícenásobné regresní analýzy

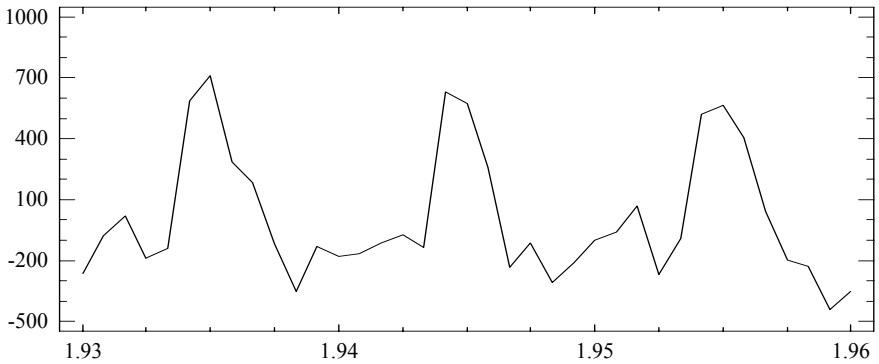
Výsledky získáme ihned po vyplnění vstupního panelu.

Graf reziduí (obr. 2.16a) zobrazíme pomocí volbou Residuals versus Row Number. Pro graf autokorelační funkce (obr. 2.16b) si nejprve uložíme rezidua do tabulkového editoru STATGRAPHICSu výběrem Residuals z nabídky a tato rezidua potom vložíme do vstupního panelu procedury pro popis časových řad **Descriptive Methods**, kde z vybereme Autocorrelation Function.

Příklad 2.12

Pokračujte v analýze předchozího příkladu a pokuste se najít nový, vhodnější model.

Autokorelaci reziduí se pokusíme odstranit tak, že trend časové řady odstraníme pomocí prvních diferencí. Graf meziměsíčních přírůstků (úbytků) počtu registrovaných nezaměstnaných absolventů gymnázií je na obr. 2.17.



Obr. 2.17: První diference řady NezGym_SR v období 1993 - 1995

Z obr. 2.17 vidíme, že řada prvních diferencí má přibližně konstantní úroveň a sezónnost. Nový regresní model má tvar

$$y_t = \beta_0 + \beta_2 D_2 + \beta_3 D_3 + \beta_4 D_4 + \ldots + \beta_{12} D_{12} + a_t.$$

Výsledky odhadu parametrů modelu jsou v tab. 2.14a.

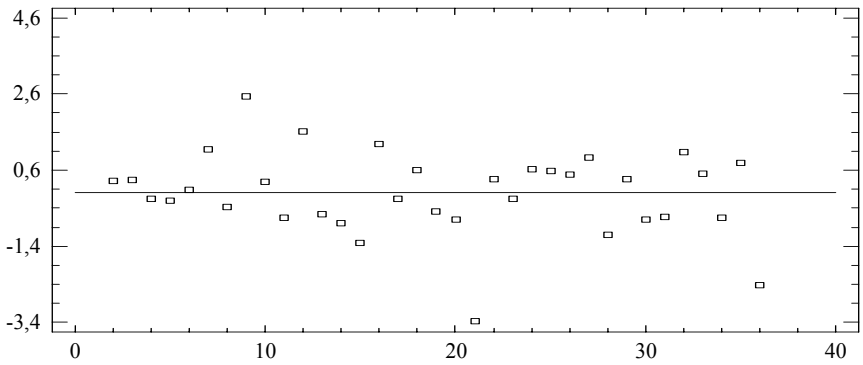
Tab.2.14a. Odhad parametrů modelu pro diferencovanou řadu

```
Multiple Regression Analysis
--------------------------------------------------------------------
Dependent variable: take(diff(NezGym_SR);36)
--------------------------------------------------------------------
                              Standard          T
Parameter   Estimate          Error          Statistic      P-Value
--------------------------------------------------------------------
CONSTANT    -140,0            71,3142        -1,96314       0,0618
D2           38,3333          92,0662         0,41637       0,6810
D3           130,667          92,0662         1,41927       0,1692
D4          -37,00            92,0662        -0,40188       0,6915
D5           20,00            92,0662         0,21723       0,8299
D6           720,333          92,0662         7,82408       0,0000
D7           755,333          92,0662         8,20424       0,0000
D8           457,333          92,0662         4,96744       0,0001
D9           137,333          92,0662         1,49168       0,1494
D10         -2,00             92,0662        -0,02172       0,9829
D11         -156,33           92,0662        -1,69805       0,1030
D12         -121,0            92,0662        -1,31427       0,2017
```

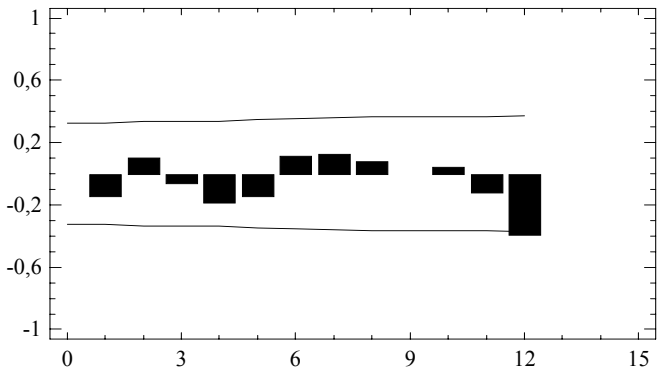
Tab. 2.14b. Analýza rozptylu modelu pro diferencovanou řadu

```
                              Analysis of Variance
--------------------------------------------------------------------------------
Source          Sum of Squares     Df     Mean Square      F-Ratio      P-Value
--------------------------------------------------------------------------------
Model            3,15968E6         11       287244,0        28,24       0,0000
Residual         233943,0          23       10171,4
--------------------------------------------------------------------------------
Total (Corr.)     3,39362E6              34

R-squared = 93,1064 percent
R-squared (adjusted for d.f.) = 89,8095 percent
Standard Error of Est. = 100,853
Mean absolute error = 63,9619
Durbin-Watson statistic = 2,15688
```

Koeficient determinace modelu (`R-squared`) je 93,1064 %. Hodnota Durbinovy-Watsonovy statistiky (`Durbin-Watson statistic`) 2,15688 signalizuje, že model nemá autokorelovanou nesystematickou složku, což potvrzuje i autokorelační funkce reziduí (obr. 2.17b). Dále lze z obr. 2.17a lze konstatovat, že nesystematická složka modelu je homoskedastická.



Obr. 2.17a: Rezidua regresního modelu pro diferencovanou řadu



Obr. 2.17b: ACF reziduí regresního modelu pro diferencovanou řadu

Z výše uvedených výsledků vyplývá, že druhý model je možné použít pro výpočet extrapolací meziměsíčních změn počtu registrovaných nezaměstnaných absolventů gymnázií, jakož i absolutního počtu registrovaných nezaměstnaných absolventů gymnázií na Slovensku v roce 1996.

Odhad modelu má tvar

$$\Delta y_t = -140{,}0 + 38{,}3333D_2 + 130{,}667D_3 - 37{,}0D_4 + 20{,}0D_5 + 720{,}333D_6 +$$
$$+ 755{,}333D_7 + 457{,}333D_8 + 137{,}333D_9 - 2{,}0D_{10} - 156{,}333D_{11} - 121{,}0D_{12}$$

Extrapolace přírůstků (úbytků) nezaměstnaných absolventů gymnázií na jednotlivé měsíce roku 1996 a jejich 95% intervaly spolehlivosti jsou v tab. 2.14c).

Tab 2.14c. Extrapolace na rok 1996 pro diferencovanou řadu

Regression Results for take(diff(NezGym_SR);36)

Row	Fitted Value	Lower 95.0% CL for Forecast	Upper 95.0% CL for Forecast
37	-140,00	-395,521	115,521
38	-101,667	-342,574	139,240
39	-9,33333	-250,240	231,574
40	-177,00	-417,907	63,907
41	-120,00	-360,907	120,907
42	580,333	339,426	821,240
43	615,333	374,426	856,240
44	317,333	76,4262	558,240
45	-2,66667	-243,574	238,240
46	-142,00	-382,907	98,9071
47	-296,333	-537,240	-55,4262
48	261,00	-501,907	-20,0929

Extrapolované hodnoty původní časové řady (počtu nezaměstnaných absolventů gymnázií) na první tři měsíce roku 1996 získáme jako

$$\hat{y}_{37} = -140 + y_{36} = -140 + 3443 = 3341 \text{ osob,}$$

$$\hat{y}_{38} = -101{,}667 + \hat{y}_{37} = -101{,}337 + 3341 = 3239{,}333\,,$$

$$\hat{y}_{39} = -9{,}333 + \hat{y}_{38} = -9{,}333 + 3239{,}333 = 3230 \text{ osob, atd.}$$

Porovnáme-li extrapolace na leden až březen roku 1996 se skutečnými hodnotami zjistíme, že počet evidovaných absolventů gymnázií v lednu roku 1996 byl 3731 osob, v únoru 3590 osob a v březnu 3473 osob. V prvním čtvrtletí 1996 tedy model podhodnocoval skutečnost.

Časovou řadu z předchozího příkladu budeme kromě zkracování ještě diferencovat, takže do prvního řádku vstupního panelu (Dependent Variable:) vložíme

TAKE(DIFF(NezGym_SR);36)

což znamená, že se z diferencované časové řady vybere pouze prvních 36 hodnot. Hodnoty umělých proměnných se nemění a časovou proměnnou t nepoužijeme.

Grafy reziduí a reziduální autokorelační funkce zobrazíme stejně jako v minulém příkladu. Pro získání předpovědí vybereme Reports z nabídky.

2.3.7 Holtovo - Wintersovo exponenciální vyrovnávání

Winters v roce 1960 rozšířil Holtovo exponenciální vyrovnávání s lineárním trendem o aditivní a multiplikativní sezónnost. Při použití Holtova-Wintersova exponenciálního vyrovnávání předpokládáme, že v úseku $t = 1, 2, ..., T$ lze časovou řadu rozložit na lokální lineární trendy s konstantní nebo multiplikativní sezónností ve tvaru

$$y_t = (\beta_0 + \beta_1 t) + S_t + a_t$$

nebo

$$y_t = (\beta_0 + \beta_1 t) \,.\, S_t \,.\, a_t,$$

kde β_0 je parametr úrovně lineárního trendu, β_1 je parametr směrnice lineárního trendu, t je časová proměnná, S_t je sezónní průměr nebo sezónní index v čase t, přičemž musí platit

1. pro sezónní průměry $\sum_{j=1}^{s} S_j = 0$,
2. pro sezónní indexy $\sum_{j=1}^{s} S_j = s$,

a_t je nesystematická složka typu bílého šumu.

Rekurentní vztahy exponenciálního vyrovnávání lineárního trendu a multiplikativní sezónnosti získáme tak, že k rovnicím (2.74) přidáme vztah pro adaptivní vyrovnávání sezónních indexů s vyrovnávací konstantou $\gamma \in <0, 1>$. Holtovo – Wintersovo exponenciální vyrovnávání je dané rekurentními vztahy:

$$\begin{aligned} \hat{U}_t &= \alpha(y_t / \hat{S}_{t-s}) + (1-\alpha)(\hat{U}_{t-1} + \hat{T}_{t-1}) \\ \hat{T}_t &= \beta(\hat{U}_t - \hat{U}_{t-1}) + (1-\beta)\hat{T}_{t-1} \\ \hat{S}_t &= \gamma(y_t / \hat{U}_t) + (1-\gamma)\hat{S}_{t-s}, \end{aligned} \tag{2.80}$$

kde $\hat{U}_t$ je odhad úrovně lineárního trendu v čase t, $\hat{T}_t$ odhad směrnice lineárního trendu v čase t, y_t hodnota časové řady v čase t, $\hat{U}_{t-1}$ odhad úrovně lineárního trendu v čase t - 1, $\hat{T}_{t-1}$ odhad směrnice lineárního trendu v čase t - 1, $\hat{S}_{t-s}$ odhad sezónního výkyvu v čase t - s, s počet sezón v roce, $\alpha \in <0, 1>$ je vyrovnávací konstanta úrovně lineárního trendu, $\beta \in <0, 1>$ je vyrovnávací konstanta směrnice lineárního trendu, $\gamma \in <0, 1>$ je vyrovnávací konstanta sezónních výkyvů.

Z první rovnice vztahu (2.80) se odhad úrovně lineárního trendu v čase t získá jako vážený aritmetický průměr sezónně očištěné hodnoty ($y_t / \hat{S}_{t-s}$) a extrapolované úrovně řady v čase t - 1. Vidíme, že dříve než známe hodnotu y_t, určujeme hodnotu lineárního trendu v čase t - 1 součtem $\hat{U}_{t-1} + \hat{T}_{t-1}$. Protože nové pozorování y_t obsahuje sezónnost, odstraníme ji vydělením hodnoty y_t hodnotou $\hat{S}_{t-s}$, tj. posledním dostupným odhadem sezónního výkyvu.

Odhad směrnice lineárního trendu v čase t je váženým aritmetickým průměrem změny úrovně lineárního trendu v čase t oproti času t - 1 a odhadu směrnice lineárního trendu v čase t - 1.

Odhad sezónního výkyvu v čase t určíme jako vážený aritmetický průměr nového pozorování y_t očištěného od odhadu úrovně lineárního trendu v čase t a posledního odhadnutého sezónního výkyvu.

Bodová předpověď a její chyba

Bodovou předpověď určitého ukazatele určíme v čase T na období $T + h$ pro multiplikativní model sezónnosti vztahem

$$\hat{y}_T(h) = (\hat{U}_T + h\hat{T}_T)\hat{S}_{T+h-s}\,. \tag{2.81}$$

Exponenciální vyrovnávání používáme obyčejně od začátku časové řady a určujeme bodové předpovědi na jedno období dopředu (h = 1), tj. v čase t - 1 na čas t, což zapisujeme pro multiplikativní model vztahem

$$\hat{y}_{t-1}(1) = (\hat{U}_{t-1} + \hat{T}_{t-1})\hat{S}_{t-s}\,, \tag{2.82}$$

a chybu předpovědi s horizontem h = 1 definujeme

$$\hat{a}_t = y_t - \hat{y}_{t-1}(1) = y_t - (\hat{U}_{t-1} + \hat{T}_{t-1})\hat{S}_{t-s}\,. \tag{2.83}$$

Interval spolehlivosti předpovědí

Podle Bowermana a O´Connella (1979) 100(1 - α/2)% předpovědní interval spolehlivosti pro y_{T+h} určíme vztahem

$$\hat{y}_T(h) \pm u_{1-\alpha/2} \cdot d_h \cdot MAE_T\,, \tag{2.84}$$

kde $\hat{y}_T(h)$ je bodová předpověď určená v čase T s horizontem $h \geq 1$, $u_{1-\alpha/2}$ je (1 - α/2)% kvantil normovaného normálního rozdělení $N(0,1)$,

$$d_h = 1{,}25\sqrt{\frac{1+\frac{\lambda}{(1+\nu)^3}((1+4\nu+5\nu^2)+2\lambda(1+3\nu)h+2\lambda^2h^2)}{1+\frac{\lambda}{(1+\nu)^3}((1+4\nu+5\nu^2)+2\lambda(1+3\nu)+2\lambda^2)}}\,, \qquad \lambda = \max(\alpha, \beta, \alpha, \beta, \gamma) \text{ a } \nu = 1 - \lambda,$$

$$MAE_T = \sum_{t=1}^{T}\frac{\left|y_t / \hat{S}_{t-s} - (\hat{U}_{t-1} + \hat{T}_{t-1})\right|}{T}\,.$$

Hodnotu MAE_T je možné s každou novou informací (novým pozorováním) obnovovat podle vztahu

$$MAE_{T+1} = \frac{T \cdot MAE_T + \left| y_{T+1} / \hat{S}_{T-s+1} - \hat{U}_T - \hat{T}_T \right|}{T+1}.$$

Volba vyrovnávacích konstant

Za nejvhodnější kombinaci vyrovnávacích konstant α, β a γ považujeme takovou kombinaci, pro kterou je rozptyl chyb předpovědí $\hat{a}_t = y_t - \hat{y}_{t-1}(1)$ pro $t = 1, 2, ..., T$ nejmenší.

Statistické softwarové produkty pracující pod Windows (STATGRAPHICS Plus, SPSS, SAS apod.) nabízí nejvhodnější kombinaci konstant automaticky.

Poznámka 1:

Rekurentní vztahy (2.80) pro lineární proces s aditivní sezónností lze zapsat analogicky jako pro multiplikativní sezónnost. Tyto vztahy se změní tak, že sezónně očištěnou hodnotu řady určíme rozdílem $(y_t - \hat{S}_{t-s})$, hodnotu časové řady očištěnou od trendu určíme rozdílem $(y_t - \hat{U}_t)$. Podobně ve vztazích (2.82) a (2.83) se násobení změní na součet a dělení na rozdíl.

Poznámka 2:

V nabídce STATGRAPHICSu Plus lze použít Holtovo - Wintersovo exponenciální vyrovnávání pouze pro multiplikativní model.

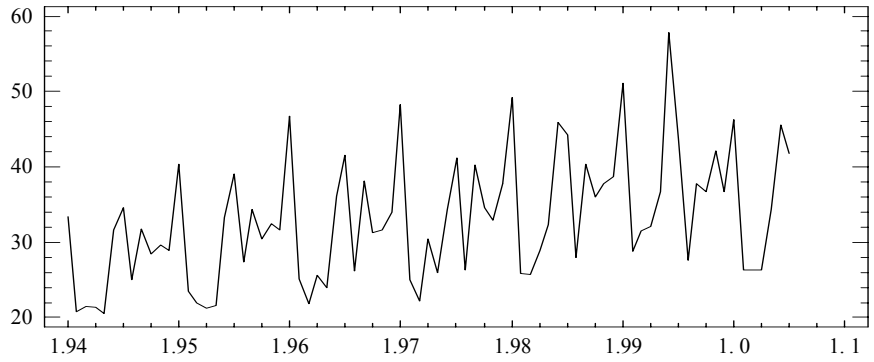
2.3.8 Wintersovo exponenciální vyrovnávání ve STATGRAPHICSu

V části 2.2.6 jsme uvedli jak používat exponenciální vyrovnávání ve STATGRAPHICSu a proto se zde o praktické stránce této metody nebudeme zmiňovat. Poznamenáme pouze, že v procedurách STATGRAPHICSu se metoda nachází v proceduře **Winter's Exp. Smoothing** ("Wintersovo exponenciální vyrovnávání").

Příklad 2.13

Analyzujte měsíční údaje o počtu nově registrovaných uchazečů o zaměstnání na Slovensku v letech 1994 až 1999 (NezNov_SR). Pomocí Wintersova modelu určete předpovědi s horizontem jeden měsíc od ledna 1994 do prosince 1999. Vyhodnoťte přesnost metody na základě charakteristik chyb předpovědí MSE, MAE, MAPE, ME a MPE v období analýzy (leden 1994 až prosinec 1999). Určete předpovědi na prvních sedm měsíců roku 2000. Ověřte přesnost předpovědí v roce 2000 pomocí relativní míry přesnosti MAPE.

Z grafu (obr. 2.18) vidíme, že vývoj časové řady má rostoucí přibližně lineární trend a výraznou sezónnost.



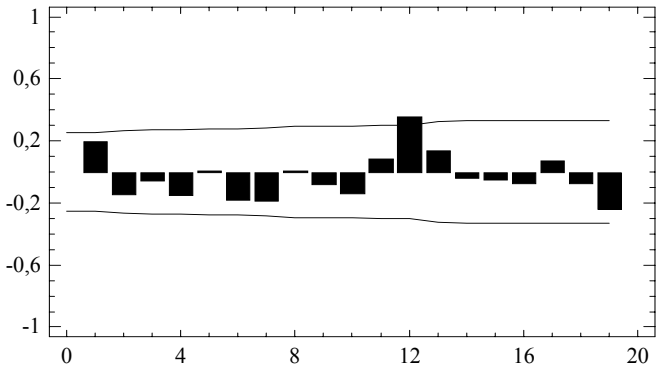
Obr. 2.18: Nově registrovaní uchazeči o zaměstnání na Slovensku od ledna 1994 do července 2000 v tis. osob

Charakteristiky chyb předpovědí Wintersova exponenciálního vyrovnávání v období od ledna 1994 do prosince 1999 jsou v tab. 2.15a. Z *MAPE* vyplývá, že předpovědi v období leden až prosinec 2000 lze určit s chybou asi 7 %.

Tab. 2.15a: Průměrné charakteristiky chyb předpovědí v čase od ledna 94 do prosince 1999

```
Data variable: NezNov_SR            T = 72            Start index = 1/94
Sampling interval = 1.0 month(s)                       Length of seasonality = 12
Forecast Summary
-------------------------------------------------------------------------------
Forecast model selected: Winter's exp. smoothing with
alpha = 0.069, beta = 0.0001, gamma = 0.2582
Number of forecasts generated: 12
Number of periods withheld for validation: 0
Estimation Period:
MSE         1,25935E7
MAE         2444,05
MAPE        7,08084
ME         -345,486
MPE        -1,83314
```

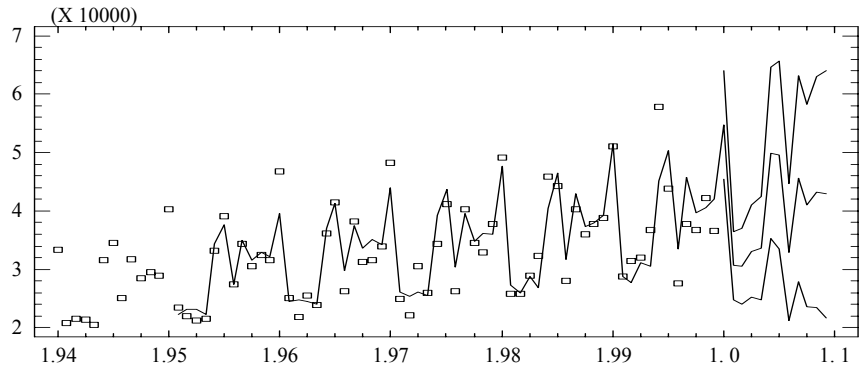
Vhodnost Wintersova modelu ověříme pomocí autokorelační funkce chyb předpovědí s horizontem jeden měsíc (obr. 2.18a).



Obr. 2.18a: Autokorelační funkce chyb předpovědí

Z obr. 2.18a plyne, že lze mít pochybnosti o kvalitě reziduí (o kvalitě chyb předpovědí) Wintersova modelu, neboť tento model pravděpodobně úplně nevystihl sezónnost (12. koeficient autokorelace je statisticky významný).

Obr. 2.18b obsahuje hodnoty nově registrovaných uchazečů o zaměstnání v období leden 1994 až prosinec 1999 a předpovědi na období leden 2000 až červenec 2000 spolu s 95% intervaly spolehlivosti.



Obr. 2.18b Počet nově registrovaných uchazečů o zaměstnání s předpověďmi

V tab 2.15b jsou uvedeny předpovědi na prvních sedm měsíců roku 2000 s 95% intervaly spolehlivosti. Tabulka dále obsahuje původní hodnoty časové řady a procentuální absolutní chyby předpovědí. Průměrná absolutní procentuální chyba předpovědí *MAPE* v roce 2000 je 15,68 %. Průměrná absolutní procentuální chyba předpovědí v analyzovaném období (v letech 1994 až 1999) je 7,08 % (viz tab. 2.15a). Chyba předpovědí se v roce 2000 se zdvojnásobila.

Tab. 2.15b. Bodové a intervalové předpovědi "ex ante"

Period	Forecast	Lower 95.0% Limit	Upper 95.0% Limit	Skutečnost	APE (%)
1/ 0	54704,2	45415,8	63992,5	46273,0	18,22
2/ 0	30641,0	24815,9	36466,2	26337,0	16,34
3/ 0	30496,1	24000,2	36992,0	26271,0	16,08
4/ 0	33097,8	25215,9	40979,7	26356,0	25,28
5/ 0	33637,0	24716,1	42557,8	34358,0	2,10
6/ 0	49921,6	35245,8	64597,4	45524,0	9,66
7/ 0	49554,0	33486,4	65621,5	41816,0	18,50

2.3.9 Cvičení

1. Pomocí autokorelační funkce otestujte sezónnost časové řady počtu nově registrovaných uchazečů o zaměstnání na Slovensku (NezNov_SR).
2. Určete sezónní indexy časové řady počtu nově registrovaných uchazečů o zaměstnání na Slovensku (NezNov_SR). Vypočítejte hodnoty sezónně očištěné časové řady a modelujte její vývoj vhodnou trendovou funkcí. Určete předpovědi této časové řady na leden až červenec 2000.

3. Odhadněte parametry regresního modelu s umělými sezónními proměnnými pro časovou řadu počtu nově registrovaných uchazečů o zaměstnání na Slovensku (NezNov_SR). Vypočítejte předpovědi na leden až červenec 2000.
4. Pomocí charakteristik MSE, MAE, MAPE, MPE a ME vyhodnoťte přesnost předpovědí z příkladů 2 a 3.
5. Na základě Wintersova modelu exponenciálního vyrovnávání určete předpovědi čtvrtletní časové řady HDP_c_CR zkrácené o jeden rok zezadu. Interpretujte charakteristiky MSE, MPE, MAE, MAPE, ME v období "Estimation period" a v období "Validation period".

3 BOXOVA-JENKINSOVA METODOLOGIE

3.1 Stochastický proces, stacionarita a autokorelační struktura

Stochastický proces je v čase uspořádaná řada náhodných veličin $\{y(s,t),\ s\in \boldsymbol{S},\ t\in \boldsymbol{T}\}$, kde $\boldsymbol{S}$ je výběrový prostor a $\boldsymbol{T}$ je indexní řada. Pro každé $t\in \boldsymbol{T}$ je $y(.,t)$ náhodná veličina definovaná na výběrovém prostoru $\boldsymbol{S}$. Pro každé $s\in \boldsymbol{S}$ je $y(s,.)$ realizace stochastického procesu definovaná na indexní řadě $\boldsymbol{T}$, tj. uspořádaná řada čísel, z nichž každé odpovídá jedné hodnotě indexní řady. *Časovou řadu* lze tedy chápat jako realizaci stochastického procesu. V dalších částech budeme předpokládat, že indexní řada je řadou celých čísel, tj. $\boldsymbol{T} = \{0, \pm1, \pm2, \ldots\}$. Pro zjednodušení budeme stochastické procesy značit jako $\{y_t,\ t = 0, \pm1, \pm2, ...\}$ resp. $\{y_t\}$ a časové řady jako y_t.

Stochastický proces je striktně stacionární, jestliže je jeho pravděpodobnostní chování invariantní vůči posunům v čase. Protože striktní stacionaritu je v praxi obtížné ověřovat, byl v analýze časových řad zaveden méně omezující pojem *slabá stacionarita* stochastických procesů.

Uvažujme stacionární proces $\{y_t,\ t = 0, \pm1, \pm2, ...\}$. Každou náhodnou veličinu lze popsat základními charakteristikami:

střední hodnotou

$$\mu_t = E(y_t), \tag{3.1}$$

rozptylem

$$\sigma_t^2 = D(y_t) = E(y_t - \mu_t)^2, \tag{3.2}$$

lineární vztah mezi dvojicemi náhodných veličin y_t a y_{t-k}, $k = ..., -1, 0, 1, ...$ charakterizuje kovarianční funkce

$$\gamma(t, t\text{-}k) = E(y_t - \mu_t)(y_{t-k} - \mu_{t-k}) \tag{3.3}$$

a korelační funkce

$$\rho(t, t-k) = \frac{\gamma(t, t-k)}{\sqrt{\sigma_t^2}\sqrt{\sigma_{t-k}^2}}. \tag{3.4}$$

Stochastický proces je *slabě stacionární* resp. *stacionární v kovariancích*, platí-li $\mu_t = \mu$, $\sigma_t^2 = \sigma^2$ pro všechna t a kovarianční a korelační funkce závisí pouze na časové vzdálenosti náhodných veličin, tj. $\gamma(t, t - k) = \gamma(t + k, t) = \gamma_k$ a $\rho(t, t - k) = \rho(t + k, t) = \rho_k$, $k = ..., -1, 0, 1, ...$.

Autokorelační funkce (ACF) podává informaci o síle lineární závislosti mezi veličinami y_t a y_{t-k}. Korelace mezi náhodnými veličinami y_t a y_{t-k} však může být způsobena jejich korelací s veličinami y_{t-1} y_{t-2} ..., y_{t-k+1}. *Parciální autokorelační funkce* (PACF) podává informaci o korelaci veličin y_t a y_{t-k} očištěnou o vliv veličin ležících mezi nimi. Parciální autokorelaci se zpožděním k vyjadřuje parciální regresní koeficient ϕ_{kk} v autoregresi k-tého řádu

$$y_t = \phi_{k1}y_{t-1} + \phi_{k2}y_{t-2} + \ldots + \phi_{kk}y_{t-k} + e_t, \tag{3.5}$$

kde veličina e_t je nekorelovaná s veličinami y_{t-j}, $j = 1, 2, \ldots, k$.

Za předpokladu stacionarity je odhadem střední hodnoty procesu μ výběrový průměr

$$\bar{y} = \frac{1}{T}\sum_{t=1}^{T} y_t \,, \tag{3.6}$$

kde T je počet hodnot časové řady. Rozptyl procesu γ_0 lze odhadnout pomocí výběrového rozptylu

$$S^2 = \frac{\sum_{t=1}^{T}(y_t - \bar{y})^2}{T} \,. \tag{3.7}$$

Odhad ρ_k je dán výběrovou autokorelací se zpožděním k

$$r_k = \frac{\sum_{t=k+1}^{T}(y_t - \bar{y})(y_{t-k} - \bar{y})}{\sum_{t=1}^{T}(y_t - \bar{y})^2}, \qquad k = 1, 2, ..., T - 1. \tag{3.8}$$

Výběrová parciální autokorelační funkce f_{kk} se odhaduje pomocí Durbinova rekurzivního vztahu

$$f_{kk} = \frac{r_k - \sum_{j=1}^{k-1} f_{k-1,j} r_{k-j}}{1 - \sum_{j=1}^{k-1} f_{k-1,j} r_j}, \tag{3.9}$$

$$f_{kj} = f_{k-1,j} - f_{kk} f_{k-1,k-j}\,, \qquad j = 1, 2,..., k - 1. \tag{3.10}$$

3.2 Stacionární procesy

3.2.1 Procesy AR

Proces AR(1)

Autoregresní proces prvního řádu lze zapsat jako

$$y_t = \phi_1 y_{t-1} + a_t, \tag{3.11}$$

kde $\{a_t\}$ je proces bílého šumu (proces s nulovou střední hodnotou, konstantním rozptylem a nulovou ACF a PACF), nebo

$$(1 - \phi_1 B) y_t = a_t, \tag{3.12}$$

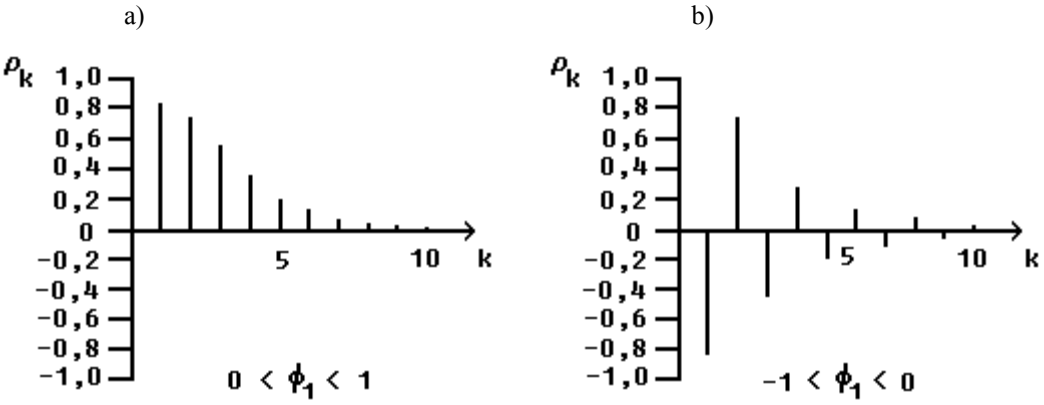
kde B je operátor zpětného posunutí, pro který platí $B^j y_t = y_{t-j}$. Jestliže $|\phi_1| < 1$, je tento proces stacionární.

Pro autokorelační funkci platí vztah

$$\rho_k = \phi_1 \rho_{k-1} = \phi_1^2 \rho_{k-2} = \ldots = \phi_1^k, \qquad k = 0, 1, 2, \ldots . \tag{3.13}$$

Parametr ϕ_1 je možné chápat jako indikátor "paměti" procesu. Čím je v absolutní hodnotě bližší jedné, tím je paměť procesu delší a naopak, čím je bližší nule, tím je paměť kratší.

Je-li roven nule, autokorelační funkce je nulová, proces nemá žádnou paměť a jedná se o proces bílého šumu. Charakteristické tvary autokorelační funkce procesu AR(1) znázorňují korelogramy na obr. 3.1.

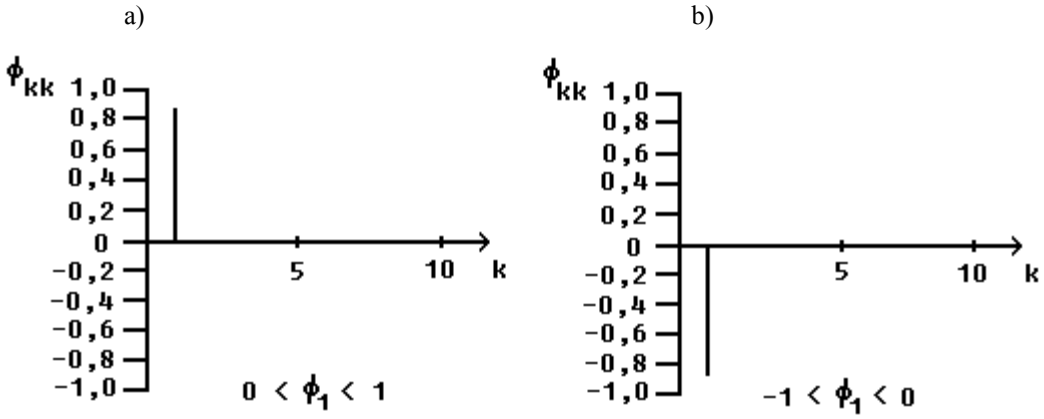


Obr. 3.1a,b: ACF procesu AR(1)

Parciální autokorelační funkce procesu AR(1) má formu

$$\phi_{kk} = \begin{cases} \rho_1 = \phi_1, & k = 1 \\ 0, & k \geq 2 \end{cases}. \tag{3.14}$$

Její tvar ukazují korelogramy na obr. 3.2.



Obr. 3.2a,b: PACF procesu AR(1)

Proces AR(2)

Autoregresní proces druhého řádu lze zapsat ve formě

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + a_t \tag{3.15}$$

nebo také jako

$$(1 - \phi_1 B - \phi_2 B^2) y_t = a_t. \tag{3.16}$$

Aby byl proces AR(2) stacionární, musí kořeny polynomiální rovnice $(1 - \phi_1 B - \phi_2 B^2) = 0$ ležet vně jednotkového kruhu.

Pro autokorelační funkci platí vztah

$$\rho_k - \phi_1\rho_{k-1} + \phi_2\rho_{k-2} = 0, \qquad k = 1, 2, \dots . \quad (3.17)$$

Potom

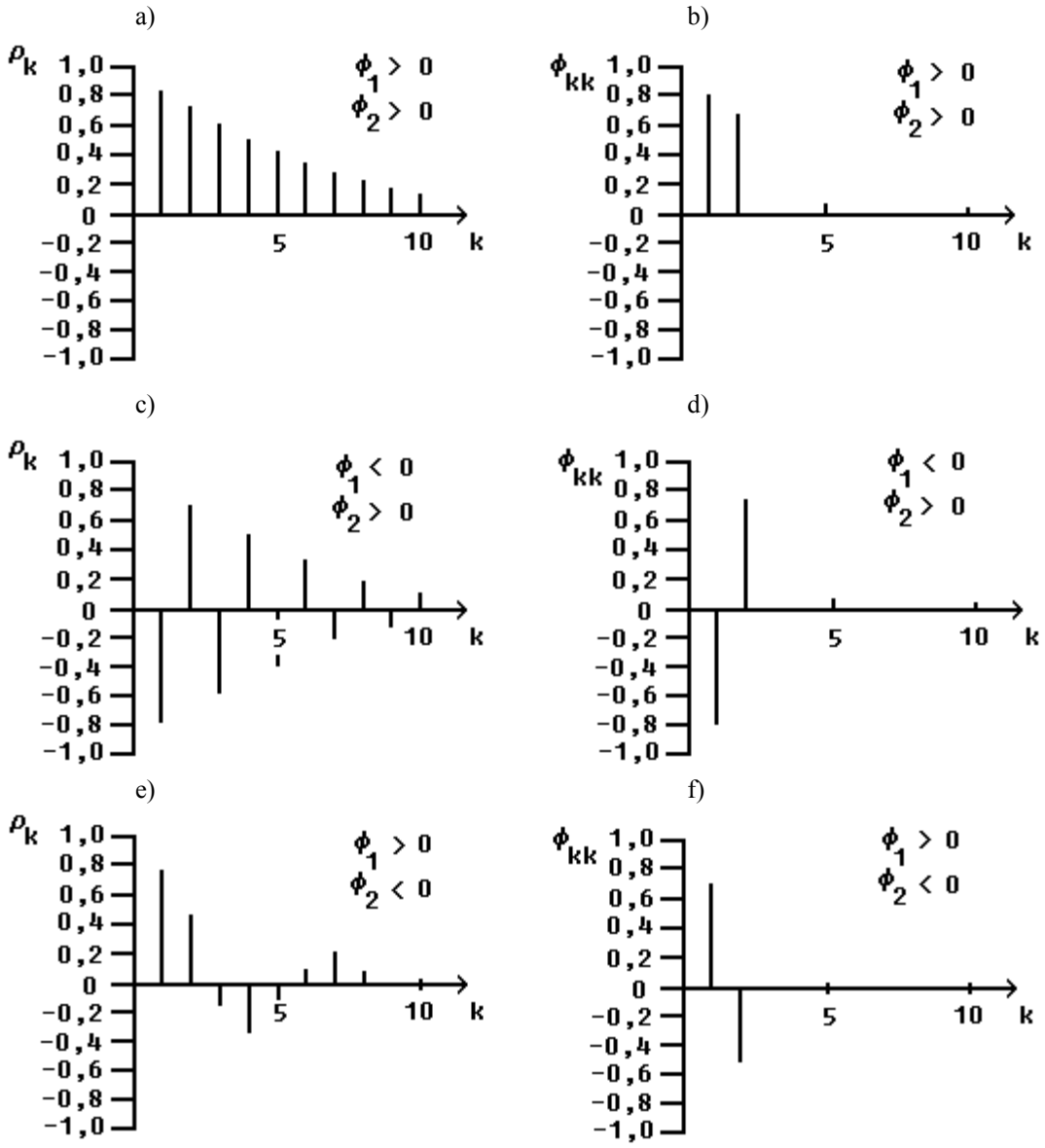
$$\rho_1 = \phi_1\rho_0 + \phi_2\rho_{-1} = \phi_1\rho_0 + \phi_2\rho_1 = \frac{\phi_1}{1-\phi_2}, \quad (3.18)$$

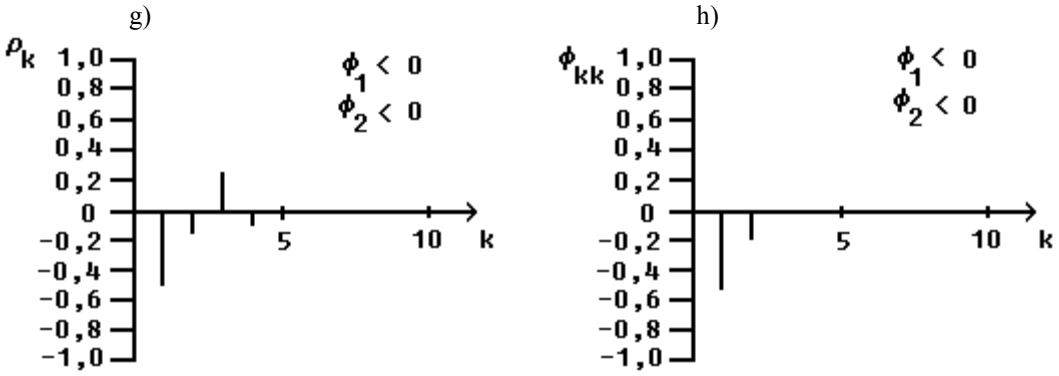
$$\rho_2 = \phi_1\rho_1 + \phi_2\rho_0 = \frac{\phi_1^2}{1-\phi_2} + \phi_2 . \quad (3.19)$$

Autokorelace v dalších zpožděních se získají rekurzívně opět na základě vzorce (3.17). Parciální autokorelační funkce nabývá hodnot

$$\phi_{11} = \frac{\phi_1}{1-\phi_2}, \ \phi_{22} = \phi_2 \text{ a } \phi_{kk} = 0, \qquad \text{pro } k = 3, 4, \dots . \quad (3.20)$$

Charakteristické tvary ACF a PACF jsou zachyceny na obr. 3.3.





Obr. 3.3a-h: ACF a PACF procesu AR(2)

Proces AR(p)

Autoregresní proces p-tého řádu AR(p) je dán vztahem

$$y_t = \phi_1 y_{t-1} + \ldots + \phi_p y_{t-p} + a_t, \tag{3.21}$$

jenž lze zapsat ve zkrácené formě jako

$$\phi_p(B) y_t = a_t, \tag{3.22}$$

kde $\phi_p(B) = (1 - \phi_1 B - \ldots - \phi_p B^p)$. Aby byl proces stacionární, musí kořeny polynomiální rovnice $\phi_p(B) = 0$ ležet vně jednotkového kruhu.

Poznámka:

V případě, že $E(y_t) = \mu \neq 0$, má model AR(p) formu $y_t = c + \phi_1 y_{t-1} + \ldots + \phi_p y_{t-p} + a_t$, kde $c = (\mu - \phi_1 \mu - \ldots - \phi_p \mu) = \mu(1 - \sum_{j=1}^{p} \phi_j)$.

3.2.2 Procesy MA

Proces MA(1)

Proces klouzavých průměrů prvního řádu má formu

$$y_t = a_t - \theta_1 a_{t-1}, \tag{3.23}$$

resp.

$$y_t = (1 - \theta_1 B) a_t, \tag{3.24}$$

kde $\{a_t\}$ je proces bílého šumu.

Tento proces je stacionární, aby byl invertibilní, tzn. aby jej bylo možné přepsat do konvergujícího procesu AR(∞), musí platit $|\theta_1| < 1$.

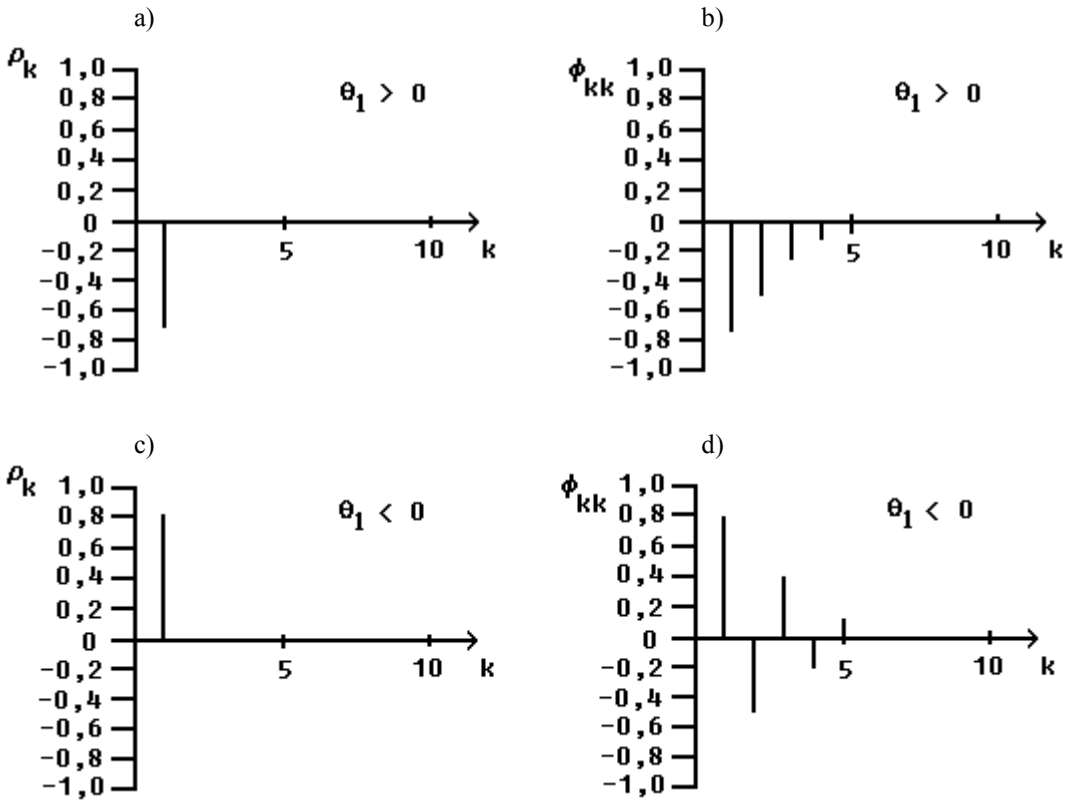
Autokorelační funkce tohoto procesu je

$$\rho_1 = \frac{-\theta_1}{1+\theta_1^2}, \quad \rho_k = 0, \qquad k = 2, 3, \ldots \tag{3.25}$$

a parciální autokorelační funkce má tvar

$$\phi_{kk} = \frac{-\theta_1^k (1-\theta_1^2)}{1-\theta_1^{2(k+1)}}, \qquad k = 1, 2, 3, \ldots \tag{3.26}$$

Autokorelační a parciální autokorelační funkce procesu MA(1) jsou znázorněny na obr. 3.4.



Obr. 3.4a-d: ACF a PACF procesu MA(1)

Proces MA(2)

Proces klouzavých průměrů druhého řádu má podobu

$$y_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2}, \tag{3.27}$$

lze jej zapsat také ve formě

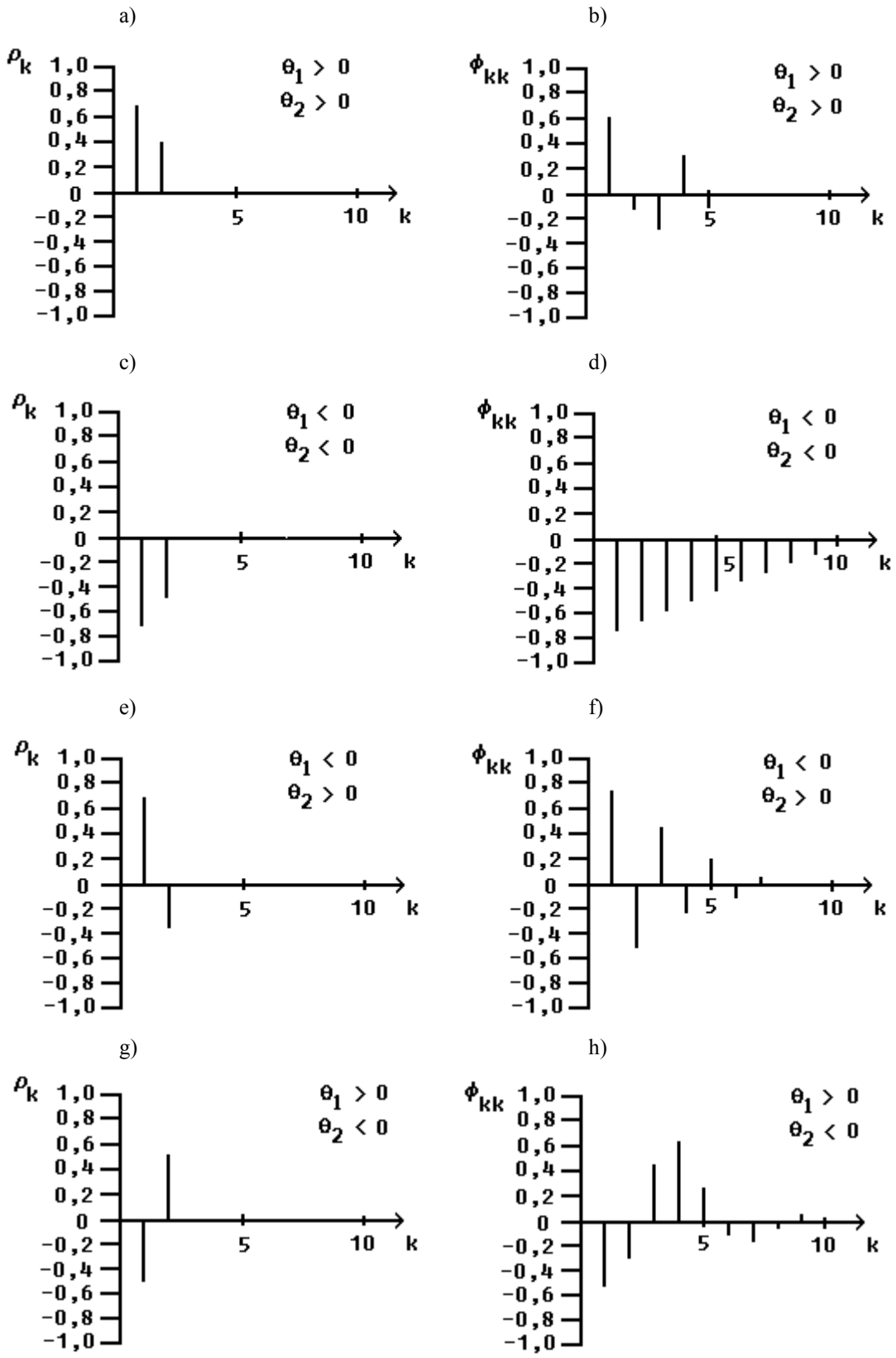
$$y_t = (1 - \theta_1 B - \theta_2 B^2) a_t. \tag{3.28}$$

Tento proces je stacionární, aby byl invertibilní, musí kořeny polynomiální rovnice $(1 - \theta_1 B - \theta_2 B^2) = 0$ ležet vně jednotkového kruhu.

Autokorelační funkce má tvar

$$\rho_k = \begin{cases} \dfrac{-\theta_1(1-\theta_2)}{1+\theta_1^2+\theta_2^2}, & k = 1, \\[2ex] \dfrac{-\theta_2}{1+\theta_1^2+\theta_2^2}, & k = 2, \\[2ex] 0, & k > 2. \end{cases} \tag{3.29}$$

Na obr. 3.5 jsou zachyceny charakteristické tvary ACF a PACF procesu MA(2).



Obr. 3.5a-h: ACF a PACF procesu MA(2)

Proces MA(q)

Proces klouzavých průměrů řádu q značíme MA(q) a lze jej zapsat ve formě

$$y_t = a_t - \theta_1 a_{t-1} - \ldots - \theta_q a_{t-q}, \tag{3.30}$$

nebo také

$$y_t = \theta_q(B) a_t, \tag{3.31}$$

kde

$$\theta_q(B) = 1 - \theta_1 B - \ldots - \theta_q B^q. \tag{3.32}$$

Proces MA(q) je stacionární. Invertibilní je tehdy, leží-li kořeny polynomu $\theta_q(B)$ vně jednotkového kruhu.

3.2.3 Procesy ARMA

Proces ARMA(1,1)

Tento proces lze vyjádřit jako

$$y_t = \phi_1 y_{t-1} + a_t - \theta_1 a_{t-1} \tag{3.33}$$

nebo také

$$(1 - \phi_1 B) y_t = (1 - \theta_1 B) a_t. \tag{3.34}$$

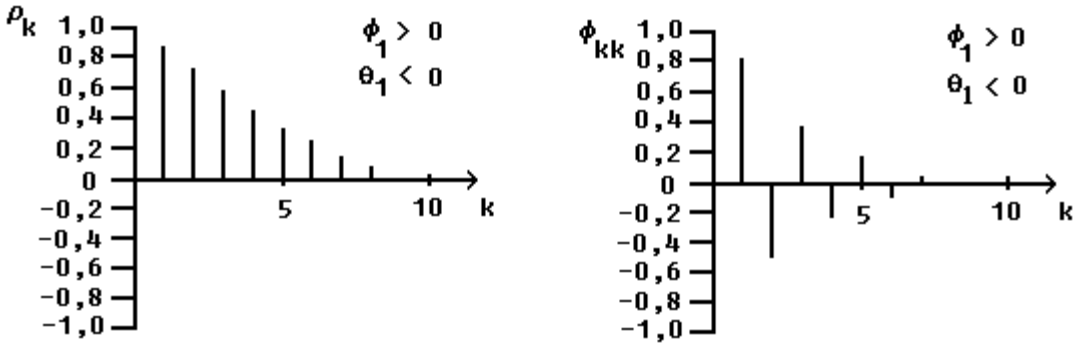
Proces je stacionární, je-li $|\phi_1| < 1$ a invertibilní, když $|\theta_1| < 1$. Jestliže $\phi_1 = 0$, proces ARMA(1,1) se redukuje na proces MA(1) a jestliže $\theta_1 = 0$, redukuje se na proces AR(1).

Autokorelační funkce má tvar

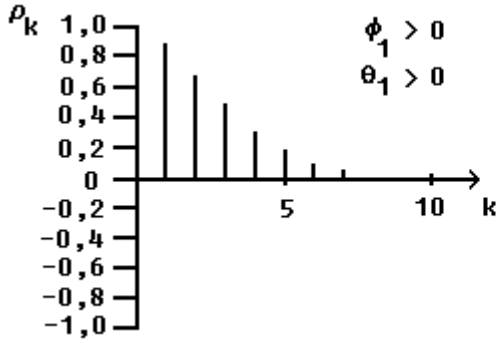
$$\rho_k = \begin{cases} \dfrac{(\phi_1 - \theta_1)(1 - \phi_1 \theta_1)}{1 - \phi_1^2 - 2\phi_1 \theta_1}, & k = 1, \\ \phi_1 \rho_{k-1}, & k \geq 2. \end{cases} \tag{3.35}$$

Obr. 3.6 znázorňují ACF a PACF procesu ARMA(1,1) pro různé hodnoty parametrů ϕ_1 a θ_1.

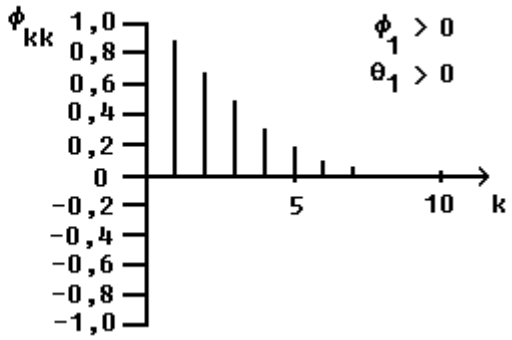
a) b)



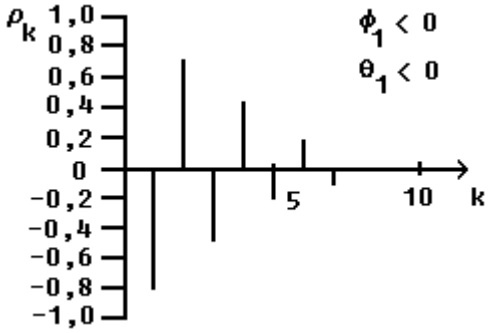
c)


ρk
1,0
0,8
0,6
0,4
0,2
0
-0,2
-0,4
-0,6
-0,8
-1,0
5
10
k
φ1 > 0
θ1 > 0

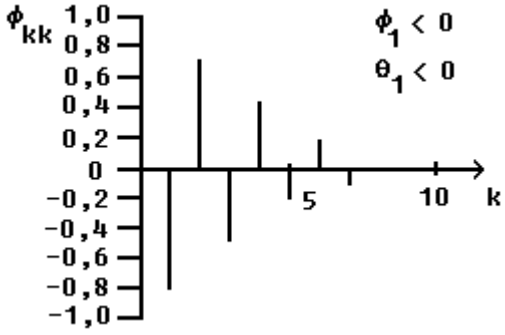
d)


φkk
1,0
0,8
0,6
0,4
0,2
0
-0,2
-0,4
-0,6
-0,8
-1,0
5
10
k
φ1 > 0
θ1 > 0

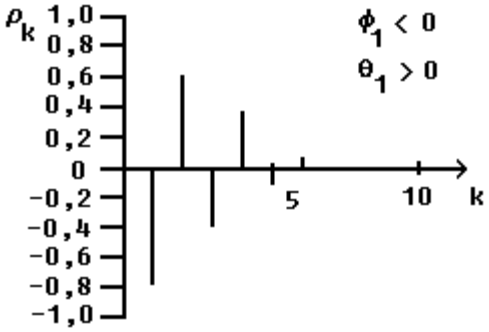
e)


ρk
1,0
0,8
0,6
0,4
0,2
0
-0,2
-0,4
-0,6
-0,8
-1,0
5
10
k
φ1 < 0
θ1 < 0

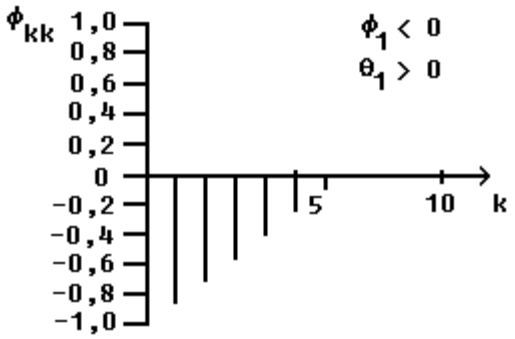
f)


φkk
1,0
0,8
0,6
0,4
0,2
0
-0,2
-0,4
-0,6
-0,8
-1,0
5
10
k
φ1 < 0
θ1 < 0

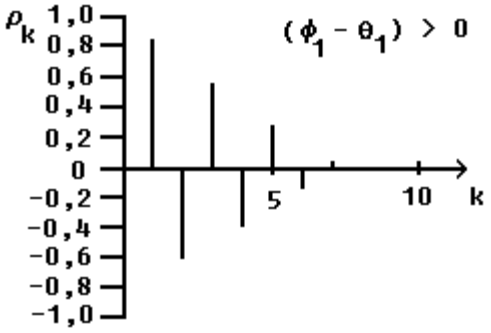
g)


ρk
1,0
0,8
0,6
0,4
0,2
0
-0,2
-0,4
-0,6
-0,8
-1,0
5
10
k
φ1 < 0
θ1 > 0

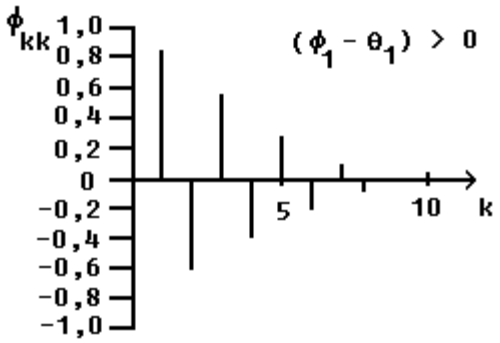
h)

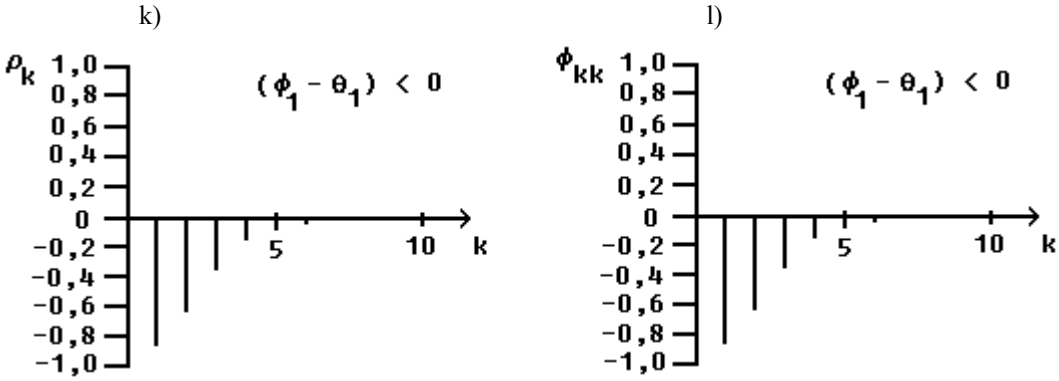

φkk
1,0
0,8
0,6
0,4
0,2
0
-0,2
-0,4
-0,6
-0,8
-1,0
5
10
k
φ1 < 0
θ1 > 0

i)


ρk
1,0
0,8
0,6
0,4
0,2
0
-0,2
-0,4
-0,6
-0,8
-1,0
5
10
k
(φ1 - θ1) > 0

j)


φkk
1,0
0,8
0,6
0,4
0,2
0
-0,2
-0,4
-0,6
-0,8
-1,0
5
10
k
(φ1 - θ1) > 0



Obr. 3.6a-l: ACF a PACF procesu ARMA(1,1)

Proces ARMA(p,q)

Proces ARMA(p,q) lze zapsat jako

$$y_t = \phi_1 y_{t-1} + \ldots + \phi_p y_{t-p} + a_t - \theta_1 a_{t-1} - \ldots - \theta_q a_{t-q} \tag{3.36}$$

nebo také jako

$$\phi_p(B) y_t = \theta_q(B) a_t, \tag{3.37}$$

kde $\phi_p(B) = 1 - \phi_1 B - \ldots - \phi_p B^p$ a $\theta_q(B) = 1 - \theta_1 B - \ldots - \theta_q B^q$. Proces ARMA($p$,$q$) je stacionární, leží-li kořeny polynomiální rovnice $\phi_p(B) = 0$ vně jednotkového kruhu a invertibilní, leží-li kořeny polynomiální rovnice $\theta_q(B) = 0$ vně jednotkového kruhu. Proces ARMA(p,q) lze přepsat do tvaru procesu MA(∞)

$$y_t = \psi(B) a_t, \tag{3.38}$$

kde $\psi(B) = \phi_p^{-1}(B)\ \theta_q(B) = (1 + \psi_1 B + \psi_2 B^2 + \ldots)$, nebo do tvaru procesu AR(∞)

$$\pi(B) y_t = a_t, \tag{3.39}$$

kde $\pi(B) = \phi_p(B)\ \theta_q^{-1}(B) = (1 + \pi_1 B + \pi_2 B^2 + \ldots)$.

3.3 Integrované procesy

3.3.1 Proces náhodné procházky

Proces náhodné procházky ("Random Walk") je možné vyjádřit jako

$$y_t = y_{t-1} + a_t, \tag{3.40}$$

lze jej zapsat také pomocí operátoru zpětného posunutí

$$(1 - B) y_t = a_t. \tag{3.41}$$

Vzhledem k tomu, že

$$(1 - B)^{-1} = (1 + B + B^2 + \ldots),$$

je možné zapsat také tento model jako

$$y_t = a_t + a_{t-1} + a_{t-2} + \ldots = \sum_{i=0}^{\infty} a_{t-i}\ . \tag{3.42}$$

V praxi je používanější modifikace procesu (3.40) zahrnující konstantu

$$y_t = c + y_{t-1} + a_t. \tag{3.43}$$

Provede-li se substituce až do $t = 0$ s počáteční hodnotou y_0, potom

$$\begin{aligned} y_t &= y_{t-1} + c + a_t \\ &= y_{t-2} + 2c + a_t + a_{t-1} \\ &\ldots \\ &= y_0 + tc + \sum_{j=1}^{t} a_j . \end{aligned} \tag{3.44}$$

Je zřejmé, že proces obsahuje *deterministický lineární trend* ($y_0 + tc$), $\sum_{j=1}^{t} a_j$ se označuje jako *stochastický trend*.

Autokorelační funkce procesu náhodné procházky závisí na čase t a s $t \to \infty$ při daném zpoždění k konverguje k jedné. Také první hodnota parciální autokorelační funkce s $t \to \infty$ konverguje k jedné, ostatní hodnoty jsou nulové.

3.3.2 Procesy ARIMA

Vykazuje-li po transformaci integrovaného procesu pomocí diference d-tého řádu výsledný proces takové autokorelace a parciální autokorelace, že jej lze vyjádřit ve formě stacionárního a invertibilního modelu ARMA(p,q), potom se původní integrovaný proces vyjádřený ve formě

$$\phi_p(B)(1 - B)^d y_t = \theta_q(B) a_t \tag{3.45}$$

nazývá *autoregresní integrovaný proces klouzavých průměrů řádu p, d, q* a označuje se jako ARIMA(p,d,q). Vlastnosti tohoto typu procesu jsou obdobné jako vlastnosti náhodné procházky. Tyto procesy se někdy nazývají pouze integrovanými procesy řádu d a označují se jako I(d).

3.4 Sezónní procesy

3.4.1 Procesy SARIMA

Myšlenka sezónního procesu je následující: jako v případě procesu ARIMA předpokládáme vzájemnou závislost mezi veličinami ... y_{t-3}, y_{t-2}, y_{t-1}, y_t, y_{t+1}, y_{t+2}, y_{t+3}, ..., a protože tento proces obsahuje ještě sezónní kolísání, lze očekávat i závislost mezi sobě odpovídajícími veličinami v jednotlivých sezónách, tj. mezi veličinami ... y_{t-2s}, y_{t-1s}, y_{ts}, y_{t+1s}, y_{t+2s}, ... , kde s je délka sezónní periody (např. u měsíčních časových řad 12, u čtvrtletních 4).

Předpokládejme, že proces obsahuje oba typy závislostí. Závislost uvnitř period je zachycena modelem ARIMA

$$\phi_p(B)(1 - B)^d y_t = \theta_q(B) b_t. \tag{3.46}$$

Proces $\{b_t\}$ obsahuje pouze sezónní závislosti a může být popsán modelem

$$\Phi_P(B^s)(1 - B^s)^D b_t = \Theta_Q(B^s) a_t, \tag{3.47}$$

kde $\Phi_P(B^s) = 1 - \Phi_1 B^s - ... - \Phi_P B^{Ps}$ a $\Theta_Q(B^s) = 1 - \Theta_1 B^s - ... - \Theta_Q B^{Qs}$. Prostřednictvím členu $(1 - B^s)$ se konstruují sezónní diference. Jestliže se procesy (3.46) a (3.47) spojí, získá se proces

$$\Phi_P(B^s)\phi_p(B)(1 - B)^d(1 - B^s)^D y_t = \theta_q(B)\Theta_Q(B^s)a_t, \tag{3.48}$$

který je označován jako SARIMA$(p,d,q)(P,D,Q)_s$, kde p je řád procesu AR, q řád procesu MA, d řád prosté diference, P řád sezónního procesu AR, Q řád sezónního procesu MA, D řád sezónní diference a s je délka sezónní periody. Podmínky stacionarity a invertibility u sezónní části jsou obdobné jako u části nesezónní.

V případě čtvrtletních časových řad lze sezónní integrovaný proces prvního řádu vyjádřit ve tvaru

$$(1 - B^4)y_t = a_t. \tag{3.49}$$

Polynomiální rovnice

$$(1 - B^4) = (1 - B)(1 + B)(1 + iB)(1 - iB) = 0 \tag{3.50}$$

má čtyři kořeny: 1, -1, i, -i. Tyto kořeny leží na jednotkové kružnici ve frekvencích 0, π, $\pi/2$ a $3\pi/2$. Frekvence $\pi/2$ a $3\pi/2$ jsou ve čtvrtletních časových řadách nerozlišitelné, proto se uvažuje pouze frekvence $\pi/2$. Frekvence 0 je nesezónní frekvencí a jednotkový kořen v této frekvenci znamená přítomnost stochastického trendu, frekvence π znamená, že každý rok obsahuje dva cykly a frekvence $\pi/2$, že každý rok obsahuje jeden cyklus. Z uvedeného vyplývá, že sezónní diferencování neznamená pouze odstraňování tzv. integrované sezónní složky, ale také odstraňování stochastického trendu, neboť sezónně integrovaný proces (3.49) zahrnuje rovněž nesezónní integrovaný proces typu náhodné procházky.

3.5 Identifikace a ověřování modelu, konstrukce předpovědí

3.5.1 Identifikace modelu

Jednou z nejtěžších úloh při výstavbě Boxových-Jenkinsových modelů je jejich identifikace. Tato úloha spočívá v rozhodnutí, jaký typ modelu vybrat. Jde o nelehkou činnost, jenž je v mnoha případech závislá na citu a zkušenosti analytika. Identifikace je přitom pouze první fází konstrukce modelů, neboť identifikovaný model je třeba ještě ověřit a upravit. Podívejme se nyní na jednotlivé kroky identifikace modelu.

Uvažujme model ARIMA(p,d,q) ve formě (3.45).

(1) Nejprve je vhodné prozkoumat graf časové řady. V mnoha případech je možné na první pohled rozpoznat přítomnost trendu. V této fázi jde především o subjektivní zhodnocení situace. Nicméně, již na základě tohoto zhodnocení je možné stacionarizovat časovou řadu či provést jiné úpravy jako je linearizace časové řady resp. její stabilizace z hlediska rozptylu pomocí logaritmické transformace. Tento typ transformace je však vhodné provádět před vlastním diferencováním časové řady. Důvodem je skutečnost, že diferencováním je možné získat i záporné hodnoty.

(2) Dalším krokem je výpočet odhadů ACF a PACF původní časové řady. Na jejich základě je možné potvrdit, že časovou řadu je třeba stacionarizovat (v případě, že hodnoty výběrové ACF a PACF v prvním zpoždění jsou velmi blízké jedné a ostatní hodnoty výběrové ACF klesají pomalu).

(3) Po stacionarizaci časové řady se použijí výběrové ACF a PACF pro identifikaci modelů AR a MA (nalezení hodnot p a q). Tato identifikace je založena na principu podobnosti výběrových ACF a PACF s teoretickými ACF a PACF. Tab. 3.1 obsahuje popis tvarů ACF a PACF pro modely AR, MA a ARMA.

Tab. 3.1: Tvary ACF a PACF modelů AR, MA a ARMA

Model	ACF	PACF
AR(p)	exponenciální a / nebo exponenciálně sinusoidní pokles	$\phi_{kk} = 0$ pro $k > p$
MA(q)	$\rho_k = 0$ pro $k > q$	Omezená exponenciálním a/nebo exponenciálně sinusoidním poklesem
ARMA(p,q)	Od zpoždění ($q - p$) pro $q > p$ exponenciální nebo exponenciálně sinusoidní pokles	Od zpoždění (p - q) pro $p > q$ omezená exponenciálním nebo exponenciálně sinusoidním poklesem

Obsahuje-li časová řada také sezónní složku, je třeba identifikovat model typu SARIMA tvaru (3.48). Princip je stejný jako v předchozím případě.

(1) Nejprve je třeba časovou řadu stacionarizovat. Pokud je to nutné, linearizuje se řada pomocí logaritmické transformace. Poté se řada diferencuje, je třeba přitom mít na paměti, že sezónní diference zahrnuje rovněž diferenci prostou. Stacionarizaci pomocí sezónní diference indikuje tvar výběrové ACF a PACF. Tyto funkce jsou charakteristické vysokými hodnotami v nesezónních a sezónních frekvencích.

(2) V další fázi se vypočítá výběrová ACF a PACF pro stacionarizovanou časovou řadu, s jejich pomocí se potom určí typ sezónních modelů SAR, SMA, či SARMA (v sezónních frekvencích mají tyto funkce statisticky významně odlišné hodnoty od nuly, tyto hodnoty však nejsou tak vysoké, aby bylo možné časovou řadu považovat za nestacionární). Po identifikaci a určení modelu sezónní složky je třeba vypočítat výběrovou ACF a PACF tentokrát pro rezidua sezónního modelu a na jejich základě posoudit, zda by bylo vhodné model doplnit o složku AR, MA či ARMA.

3.5.2 Diagnostická kontrola modelu

Testování nesystematické složky

V modelu ARIMA(p,d,q) typu (3.45) předpokládáme, že $\{a_t\}$ je proces bílého šumu. Základní empirická diagnostická kontrola spočívá v posouzení reziduí

$$\hat{a}_t = [\hat{\theta}_q(B)]^{-1}\hat{\phi}_p(B)y_t \ , \tag{3.51}$$

kde $\hat{\phi}_p(B) = 1 - \hat{\phi}_1 B - ... - \hat{\phi}_p B^p$ a $\hat{\theta}_q(B) = 1 - \hat{\theta}_1 B - ... - \hat{\theta}_q B^q$.

Autokorelaci nesystematické složky lze testovat pomocí výběrové autokorelační funkce

$$r_k = \frac{\sum_t \hat{a}_t \hat{a}_{t-k}}{\sum_t \hat{a}_t^2} \ . \tag{3.52}$$

Není-li nesystematická složka autokorelovaná, měly by hodnoty této funkce ležet uvnitř intervalu $\pm 2/\sqrt{T}$ (95% interval spolehlivosti).

Další možností jak zjistit, zda nesystematická složka není autokorelovaná, je použití *portmanteau testu*, který navrhli Box a Pierce. Testuje se hypotéza H_0: $\rho_1 = \rho_2 = ... = \rho_K = 0$ proti hypotéze H_1: non H_0, kde ρ_k, $k = 1, ..., K$ jsou autokorelace nesystematické složky modelu pro zpoždění k. Je-li model ARIMA správně zkonstruovaný, potom má statistika

$$Q = T\sum_{k=1}^{K} \hat{r}_k^2 , \tag{3.53}$$

pro vysoké T a K přibližně rozdělení χ^2 s $(K - p - q)$ stupni volnosti.

Testování parametrů modelu

Nevykazuje-li nesystematická složka autokorelaci, je vhodné ještě otestovat jednotlivé parametry v modelu, tj. parametry μ, ϕ_i, $i = 1, 2, \ldots, p$ a θ_i, $i = 1, 2, \ldots, q$. Testování se provádí na základě t-testů resp. pomocí statistik

$$t_{\mu_i} = \frac{\hat{\mu}_i}{\hat{S}_\mu},\ t_{\phi_i} = \frac{\hat{\phi}_i}{\hat{S}_{\phi_i}},\ i = 1, 2, \ldots, p \text{ a } t_{\theta_i} = \frac{\hat{\theta}_i}{\hat{S}_{\theta_i}}\ i = 1, 2, \ldots, q, \tag{3.54}$$

kde $\hat{S}_\mu, \hat{S}_{\phi_i}, \hat{S}_{\theta_i}$ jsou odhady směrodatných chyb odhadů jednotlivých parametrů modelu ARIMA.

3.5.3 Výpočet předpovědí

Při výpočtu předpovědí budeme vycházet ze skutečnosti, že předpověď $y_T(h)$ je dána podmíněnou střední hodnotou $E(y_{T+h}|y_T, y_{T-1}, y_{T-2}, \ldots)$. Předpokládejme model ARMA(p, q) ve formě (3.36), tj.

$$y_t = \phi_1 y_{t-1} + ... + \phi_p y_{t-p} + a_t - \theta_1 a_{t-1} - ... - \theta_q a_{t-q}. \tag{3.55}$$

Pro $t = T + h$ dostaneme

$$y_{T+h} = \phi_1 y_{T+h-1} + ... + \phi_p y_{T+h-p} + a_{T+h} - \theta_1 a_{T+h-1} - ... - \theta_q a_{T+h-q}. \tag{3.56}$$

Budeme-li uvažovat podmíněné střední hodnoty v čase T

$$y_T(h) = \phi_1 y_T(h-1) + ... + \phi_p y_T(h-p) + a_T(h) - \theta_1 a_T(h-1) - ... - \theta_q a_T(h-q), \tag{3.57}$$

kde

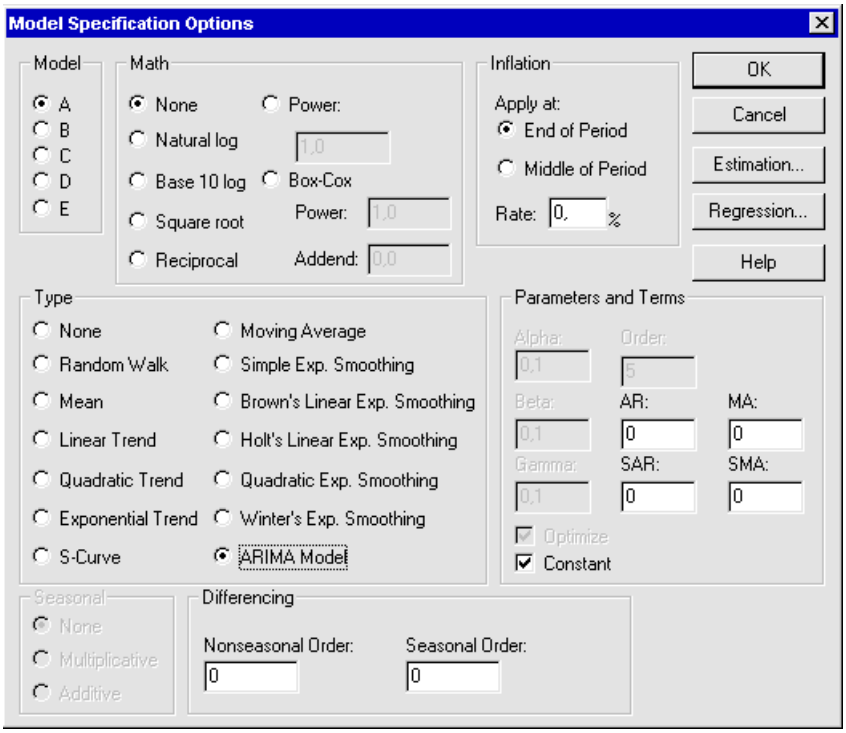
$$\begin{aligned} y_T(j) &= E(y_{T+j}|y_T, y_{T-1}, \ldots), && j \geq 1, \\ y_T(j) &= y_{T+j}, j \leq 0, \\ a_T(j) &= 0, j \geq 1, \\ a_T(j) &= y_{T+j} - y_{T+j-1}(1) = a_{T+j}, && j \leq 0. \end{aligned}$$

3.5.4 Boxova-Jenkinsova metodologie ve STATGRAPHICSu

Boxova-Jenkinsova metodologie je ve STATGRAPHICSu obsažena v nabídce

Special
⇒ **Time-Series Analysis**
⇒ **Forecasting**.

Po vyplnění vstupního panelu (obr. 2.1) vybereme Boxovy-Jenkinsovy modely tak, že v doplňkovém panelu ***Model Specifications Options*** (obr. 3.7), kde v části **Type** zvolíme ARIMA Model.



Obr. 3.7: Doplňkový panel Model Specifications Options

Volbou této položky se automaticky zaktivní v části **Parameters and Terms** a **Differencing** políčka určená pro specifikaci typu ARIMA modelu. V **Parameters and Terms** jsou to pro nesezónní časové řady políčka AR (řád autoregresního procesu AR) a MA (řád procesu klouzavých průměrů MA). Modely sezónních časových řad volíme kombinací políček AR, MA s SAR (řád sezónního autoregresního procesu SAR) a SMA (řád sezónního procesu klouzavých průměrů SMA), podmínkou je však specifikace délky sezónnosti v políčku Seasonality ve vstupním panelu. V políčku Constant zadáváme požadavek na zahrnutí konstanty do modelu.

Stacionarizující transformace časových řad provádíme pomocí položek v částech **Differencing** a **Math**.

V části **Differencing** se volí řád nesezónního integrovaného procesu (Nonseasonal Order) nebo sezónního integrovaného procesu (Seasonal Order).

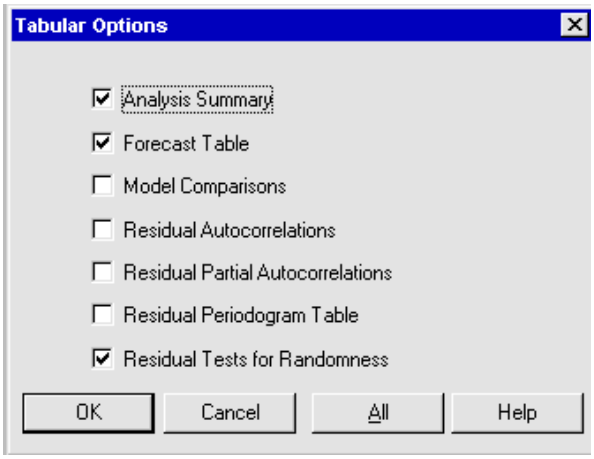
Poznámka:

Pro provedení počáteční analýzy je třeba zadat do všech zmíněných políček hodnotu 0.

V **Math** lze vybírat z následujících transformací časové řady

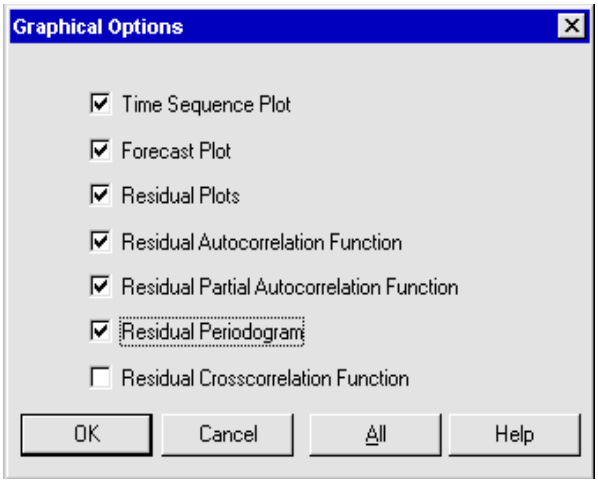
None	- žádná,
Natural log	- logaritmická transformace se základem „e“,
Base 10 log	- logaritmická transformace se základem „10“,
Square root	- odmocninová transformace,
Reciprocal	- posloupnost převrácených hodnot,
Power	- transformace ve formě *r*-tých mocnin
Box-Cox	- Boxova-Coxova transformace.

K identifikaci, ověřování ARIMA modelů a ke konstrukci předpovědí slouží ve STATGRAPHICSu nabídky výpočetních metod a grafů.



Odhad modelu a výstupní tabulka
Předpovědi
Porovnání modelů
Reziduální autokorelační funkce (ACF)
Rezid. parciální autokorelační funkce (PACF)
Reziduální periodogram
Test nahodilosti reziduí

Obr. 3.8a: Nabídka výpočetních metod



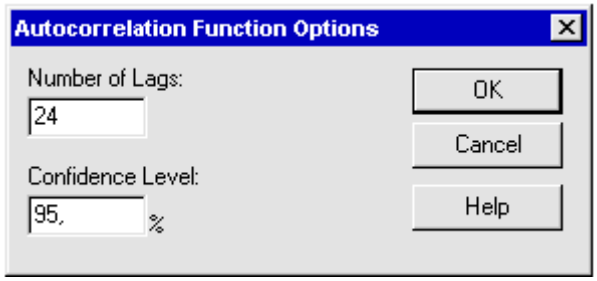
Graf časové řady
Graf časové řady s předpověďmi
Graf reziduí
Graf reziduální autokorelační funkce
Graf rezid. parciální autokorelační funkce
Graf reziduálního periodogramu
Graf reziduální vzájemné korelační funkce

Obr. 3.8b: Nabídka grafů

Z nabídky grafů (obr. 3.8b) zvolíme graf časové řady (Time Sequence Plot), graf reziduí (Residual Plots), graf autokorelační funkce (Residual Autocorrelation Function), parciální autokorelační funkce (Residual Partial Autocorrelation Function) a periodogram (Residual Periodogram). Z nabídky výpočetních metod (obr. č. 3.8a) vybereme (kromě Analysis Summary) testy nahodilosti reziduí (Residual Tests for Randomness) mezi nimiž je obsažen i Portmanteau test, zde označovaný jako Boxův-Pierceův test.

Při výpočtu předpovědí zobrazíme graf předpovědí (Forecast Plot) a graf původní časové řady s vyrovnanými hodnotami a předpověďmi (Time Sequence Plot) výběrem z nabídky a tabulku hodnot předpovědí (Forecast Table) z nabídky .

V případě nabídek tabulek a grafů autokorelační funkce, parciální autokorelační funkce a předpovědí nebo testu nahodilosti reziduí je někdy třeba pomocí doplňkového panelu (např. obr. 3.9) změnit nastavený počet zpoždění (Number of Lags) nebo šířku intervalů spolehlivosti (Confidence Level).

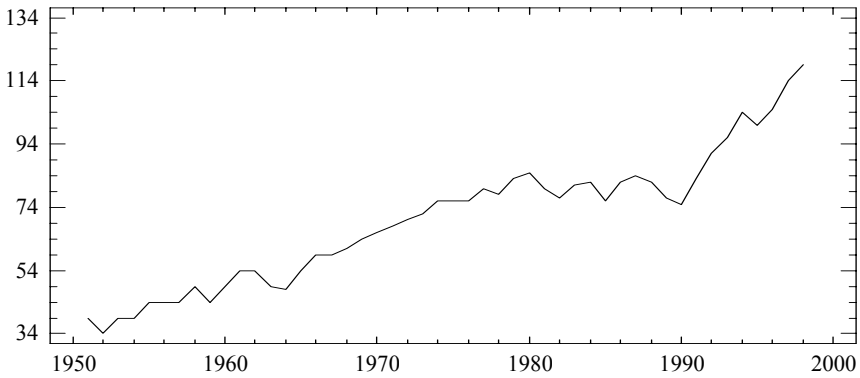


Obr. 3.9: Doplňkový panel ACF

Počet zpoždění pro výpočet tabulky a zobrazení grafu autokorelační a parciální autokorelační funkce se volí maximálně $T/4$. Pro Boxův-Pierceův test, obsažený v položce Residual Tests for Randomness, se počet zpoždění stanoví podle čísla $\sqrt{T}$.

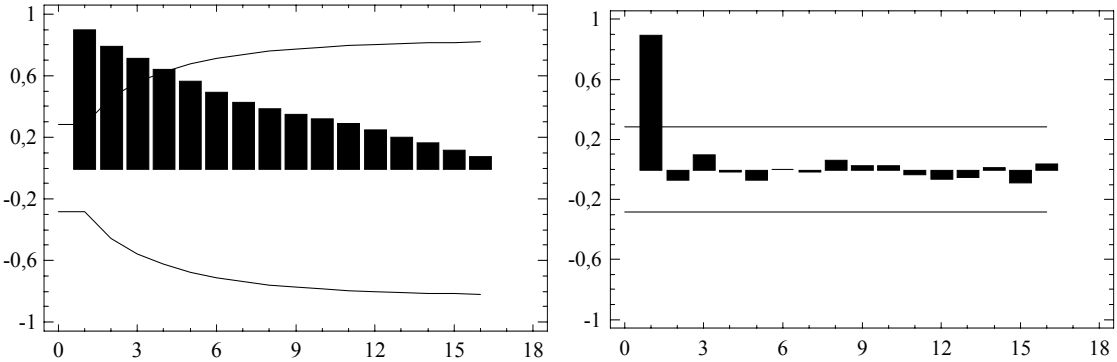
Příklad 3.1

Pomocí Boxovy-Jenkinsovy metodologie najděte model roční časové řady hrubého domácího produktu Argentiny (HDP_ARG), která je k dispozici ve formě bazických indexů od roku 1951 do roku 1998 (1995 = 100). Na základě zvoleného modelu vypočítejte předpovědi této časové řady na 4 roky.

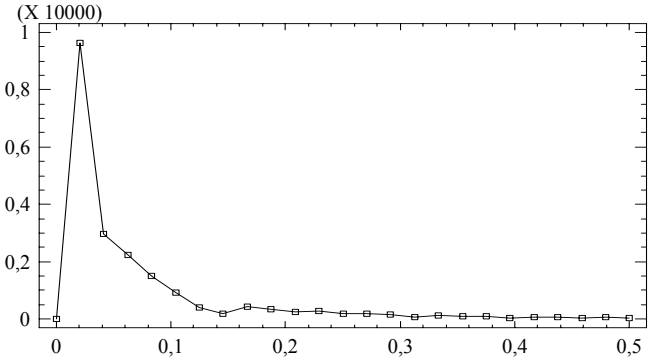


Obr. 3.10: Časová řada hrubého domácího produktu Argentiny

Průběh časové řady zachycuje obr. 3.10. Je zřejmé, že časová řada je nestacionární, což potvrzuje také tvar ACF a PACF (obr. 3.10a,b) a periodogram (obr. 3.10c).



Obr. 3.10a,b: ACF a PACF původní časové řady



Obr. 3.10c: Periodogram původní časové řady

Hodnoty ACF klesají pomalu a první hodnota u ACF i PACF je blízká jedné, periodogram má významný vrchol v nulové frekvenci, takže lze předpokládat, že časová řada je typu I(1). Budeme ji tedy stacionarizovat prvními diferencemi.

Tab. 3.2a. Výstupní tabulka modelu ARIMA(0,1,0)$_c$

```
Nonseasonal differencing of order: 1
Forecast model selected: ARIMA(0,1,0) with constant
Number of forecasts generated: 4  Number of periods withheld for validation: 0

            Estimation      Validation
Statistic   Period          Period
--------------------------------------------
MSE         16,1702
MAE         3,32096
MAPE        5,08991
ME         -1,51179E-15
MPE        -0,482326

                            ARIMA Model Summary
Parameter           Estimate        Stnd. Error     t               P-value
--------------------------------------------------------------------------------
Mean                1,70213         0,586555        2,90191         0,005675
Constant            1,70213
--------------------------------------------------------------------------------
Estimated white noise variance = 16,1702 with 46 degrees of freedom
Estimated white noise standard deviation = 4,02122

Tests for Randomness of residuals
Box-Pierce Test
     Test based on first 15 autocorrelations
     Large sample test statistic = 7,07587
     P-value = 0,95551
```

Tab. 3.2a obsahuje charakteristiky odhadnutého modelu ARIMA(0,1,0)$_c$. Tento model má formu

$$(1 - B)y_t = 1{,}70213 + a_t,$$

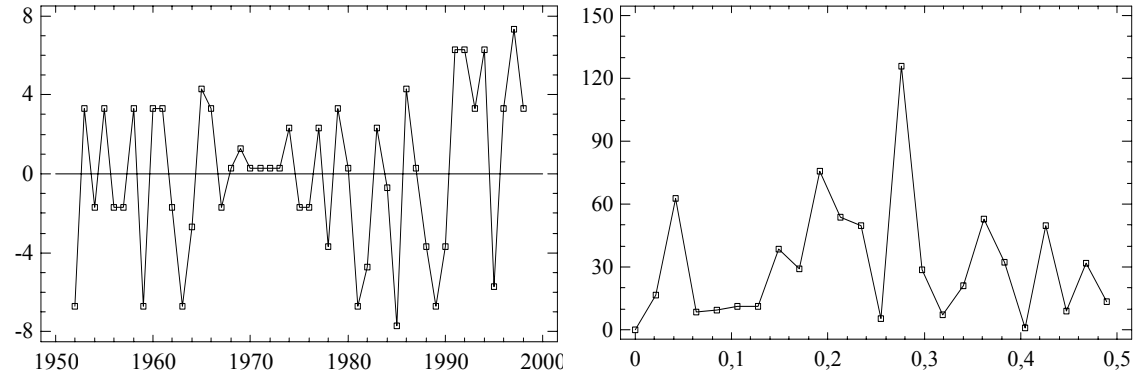
lze jej vyjádřit také jako

$$y_t = 1{,}70213 + y_{t-1} + a_t.$$

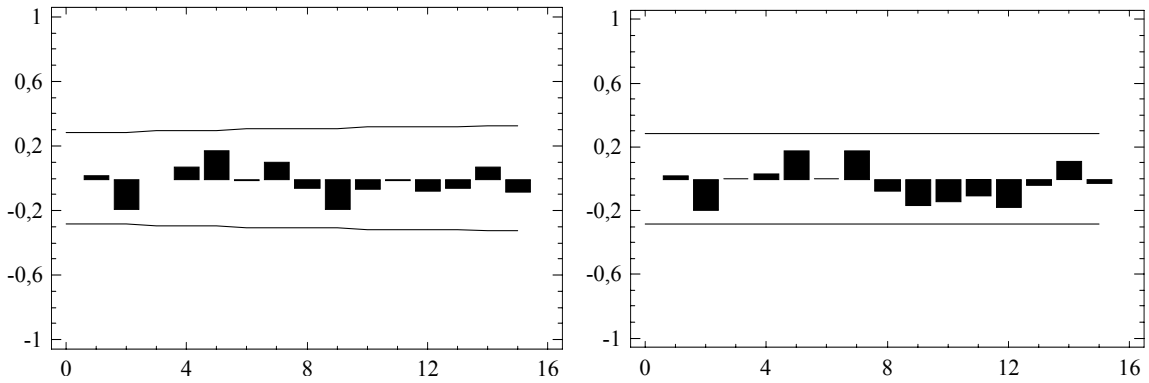
Současně otestujeme, zda je konstanta $c = \mu$ statisticky významná, tj. různá od nuly. Výsledek t-testu pro střední hodnotu μ je rovněž uveden v tab. 3.2a, porovnáme-li hodnotu "P-value" (0,005675) s hladinou významnosti α (0,05) prokážeme, že střední hodnota a konstanta jsou různé od nuly.

Autokorelaci nesystematické složky testujeme pomocí Boxova-Pierceova testu v dolní části tab. 3.2a. Vysoká hodnota "P-value" tohoto testu (0,95551) indikuje, že nesystematická složka je typu bílého šumu.

Tento závěr potvrzuje i graf reziduí a reziduální periodogram (obr. 3.10d,e), stejně jako reziduální ACF a PACF tohoto modelu (obr. 3.15f,g), jejichž hodnoty leží uvnitř tolerančních mezí.



Obr. 3.10d,e: Rezidua a reziduální periodogram modelu ARIMA(0,1,0)$_c$



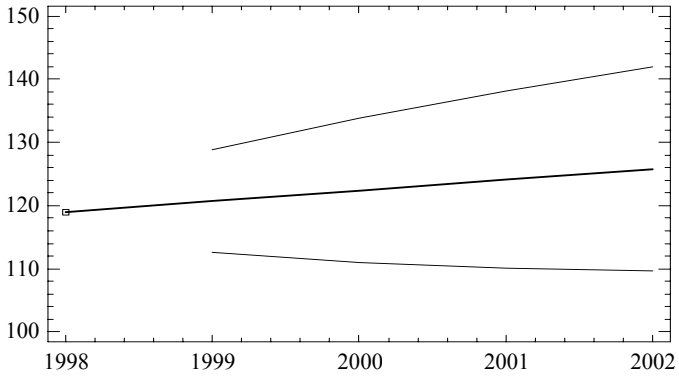
Obr. 3.10f,g: Reziduální ACF a PACF modelu ARIMA(0,1,0)$_c$

Identifikovali jsme model této časové řady ve tvaru náhodné procházky. Použijeme-li jej i pro výpočet předpovědí, získáme bodové a intervalové předpovědi na období 1999 - 2002 (tab. 3.2b).

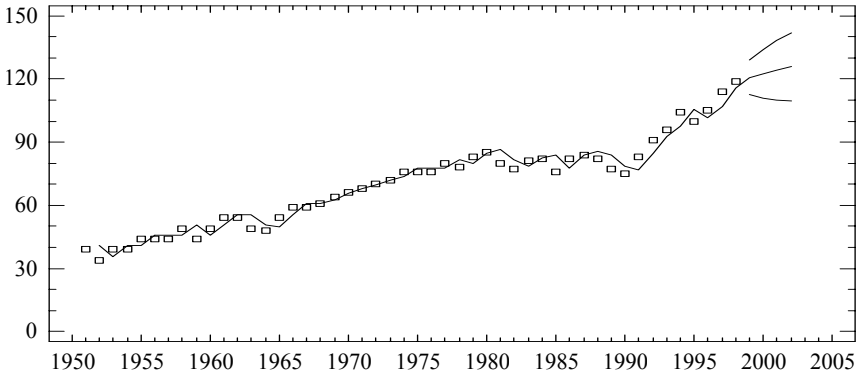
Tab. 3.2b. Předpovědi modelu ARIMA(0,1,0)$_c$

```
Forecast Table for HDP_Arg
Model: ARIMA(0,1,0) with constant

                                    Lower 95,0%         Upper 95,0%
Period          Forecast            Limit               Limit
-------------------------------------------------------------------
1999            120,702             112,608             128,796
2000            122,404             110,957             133,851
2001            124,106             110,087             138,126
2002            125,809             109,62              141,997
```

Graf předpovědí a graf původní časové řady s vyrovnanými hodnotami a předpověďmi obsahují obr. 3.10h a 3.10i.



Obr. 3.10h: Předpovědi modelu ARIMA(0,1,0)$_c$

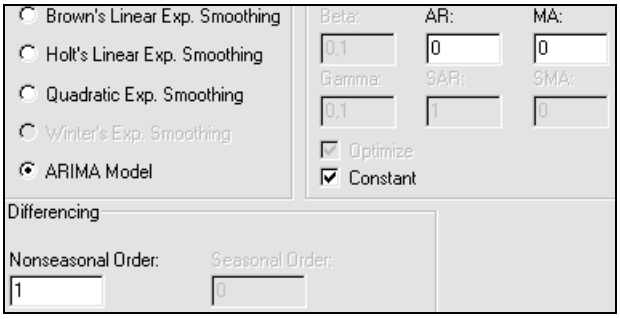


Obr. 3.10i: Původní časová řada s předpověďmi

1) Ve vstupním panelu (obr. 2.1) vložíme do řádku Data časovou řadu HDP_ARG, v části **Sampling Interval** vybereme Year(s) a v políčku Starting At přepíšeme nabízenou hodnotu na 1951. Počet předpovědí 4 zadáváme do políčka Number of Forecasts.
2) V doplňkovém panelu ***Model Specifications Options*** (obr. 3.7) v části **Type** zvolíme ARIMA Model a do všech políček v **Parameters and Terms** a **Differencing** zadáme hodnotu 0.

3) Z vybereme graf původní časové řady (obr. 3.10), graf autokorelační funkce, parciální autokorelační funkce (obr. 3.10a,b) a periodogram (obr. 3.10c) a z navíc testy nahodilosti reziduí.

4) Podle grafů ACF, PACF a periodogramu původní časové řady zvolíme pro transformaci časové řady nesezónní diferenci přepsáním nuly na jedničku v políčku Nonseasonal Order v části **Differencing**. Odhadnutý model ARIMA$(0,1,0)_c$ (tab 3.2a) se objeví automaticky ve výsledkovém okně, současně se přepočítá Boxův-Pierceův test a podle vypočítaných reziduí se překreslí všechny grafy (obr. 3.10d-g).



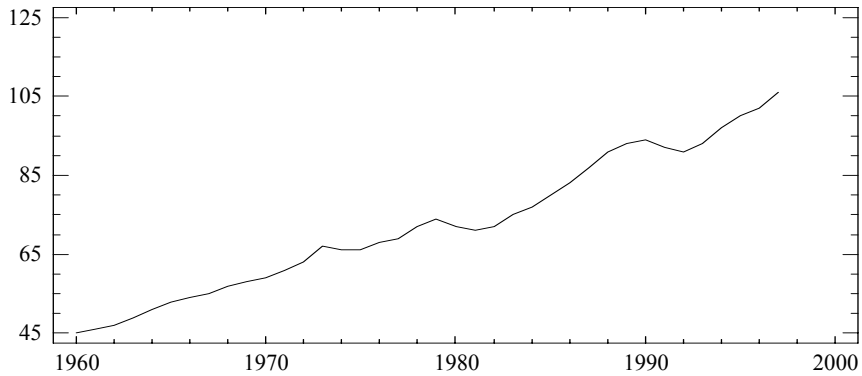
Obr. 3.10j: Výřez vyplněného doplňkového panelu Model Specifications Options pro model ARIMA$(0,1,0)_c$

5) Na základě zvoleného modelu ARIMA$(0,1,0)_c$ vypočítáme předpovědi (tab. 3.3b). Zobrazíme graf předpovědí (obr. 3.10h) a graf původní časové řady s vyrovnanými hodnotami a předpověďmi (obr. 3.10i).

Příklad 3.2

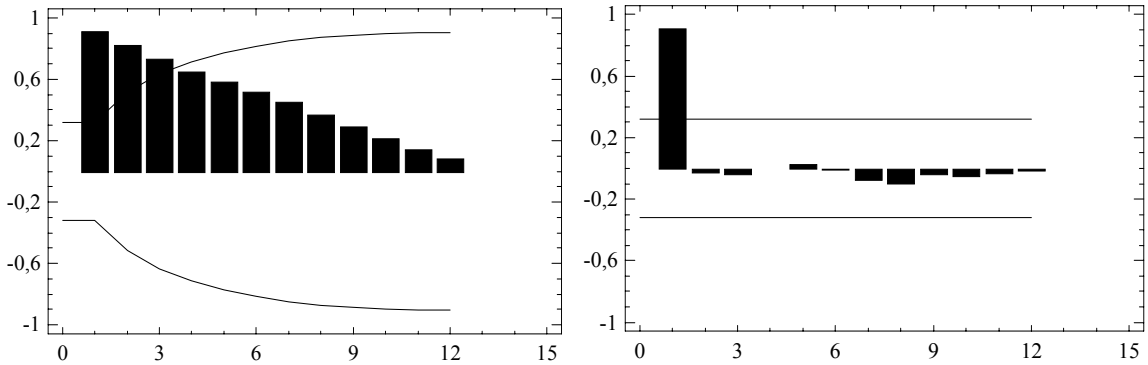
Pro roční časovou řadu hrubého domácího produktu Velké Británie (HDP_VB), která je k dispozici ve formě bazických indexů od roku 1960 do roku 1997 (1995 = 100) najděte vhodný ARIMA model a na jeho základě vypočítejte předpovědi této časové řady na 3 roky.

Průběh časové řady je zachycen na obr. 3.11. Z grafu je zřejmé, že tato časová řada je nestacionární.

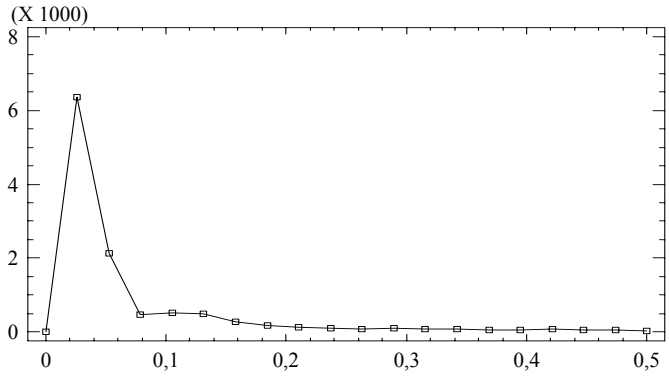


Obr. 3.11: Časová řada hrubého domácího produktu Velké Británie

Nestacionaritu potvrzuje také tvar ACF a PACF (obr. 3.11a,b) a periodogram (obr. 3.11c).



Obr. 3.11a,b: ACF a PACF původní časové řady



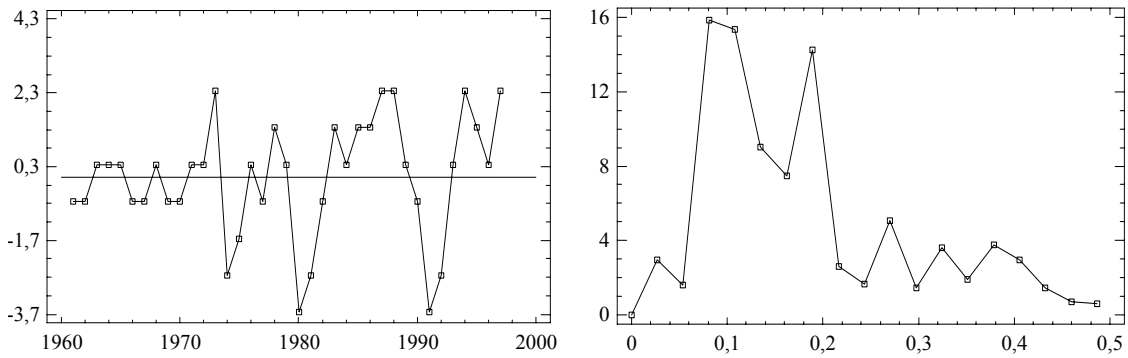
Obr. 3.11c: Periodogram původní časové řady

Protože hodnoty ACF klesají pomalu, první hodnota u ACF i PACF je blízká jedné a periodogram má významný vrchol v nulové frekvenci, lze předpokládat, že časová řada je typu I(1), a proto ji budeme stacionarizovat prvními diferencemi.

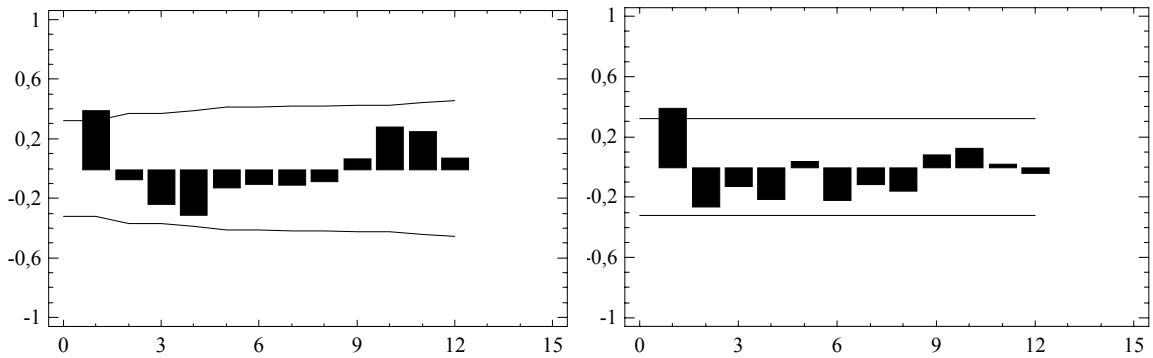
Tab. 3.3a. Výstupní tabulka modelu ARIMA(0,1,0)$_c$

```
Nonseasonal differencing of order: 1
Forecast model selected: ARIMA(0,1,0) with constant
Number of forecasts generated: 3  Number of periods withheld for validation: 0

            Estimation      Validation
Statistic   Period          Period
--------------------------------------------
MSE         2,56757
MAE         1,22863
MAPE        1,62859
ME          3,07262E-15
MPE         -0,106827

                            ARIMA Model Summary
Parameter           Estimate        Stnd. Error     t               P-value
-----------------------------------------------------------------------------
Mean                1,64865         0,263427        6,25847         0,000000
Constant            1,64865
-----------------------------------------------------------------------------
Estimated white noise variance = 2,56757 with 36 degrees of freedom
Estimated white noise standard deviation = 1,60236

Tests for Randomness of residuals
Box-Pierce Test
     Test based on first 12 autocorrelations
     Large sample test statistic = 18,8926
     P-value = 0,0911537
```

Rezidua a reziduální periodogram časové řady po I. diferenci zobrazuje obr. 3.11d,e, tvar reziduální ACF a PACF obr. 3.11f,g.



Obr. 3.11d,e: Rezidua a reziduální periodogram modelu ARIMA(0,1,0)$_c$

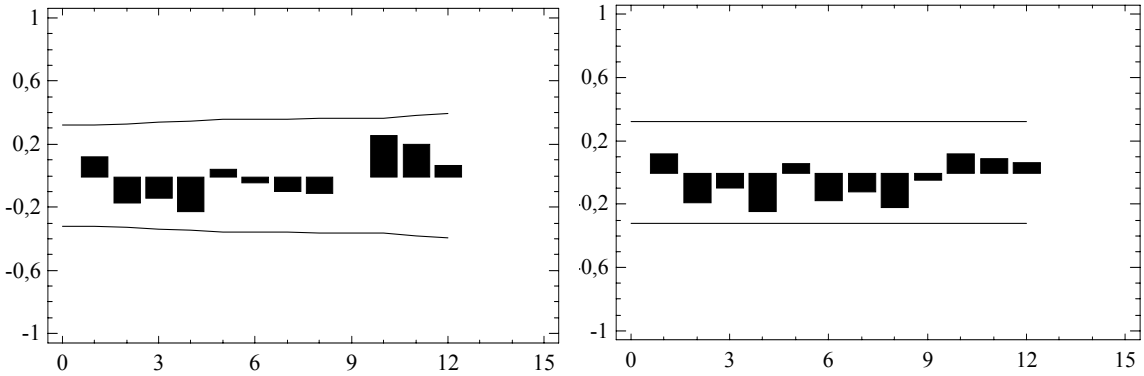


Obr. 3.11f,g: Reziduální ACF a PACF modelu ARIMA(0,1,0)$_c$

Reziduální ACF i PACF mají první hodnoty statisticky významně odlišné od nuly. Nelze proto jednoznačně určit, zda je model třeba rozšířit o část AR(1) či MA(1). Odhadneme tedy nejprve model ARIMA(1,1,0)$_c$. Výsledky jsou v tab. 3.3b.

Tab. 3.3b. Výstupní tabulka modelu ARIMA(1,1,0)$_c$

```
Nonseasonal differencing of order: 1
Forecast model selected: ARIMA(1,1,0) with constant
Number of forecasts generated: 3  Number of periods withheld for validation: 0

            Estimation      Validation
Statistic   Period          Period
--------------------------------------------
MSE         2,21185
MAE         1,09575
MAPE        1,46969
ME          0,0162096
MPE         -0,0476283

                            ARIMA Model Summary
Parameter           Estimate        Stnd. Error     t               P-value
-----------------------------------------------------------------------------
AR(1)               0,407307        0,159438        2,55464         0,015137
Mean                1,66965         0,394347        4,23396         0,000158
Constant            0,989589
-----------------------------------------------------------------------------
Estimated white noise variance = 2,21333 with 35 degrees of freedom
Estimated white noise standard deviation = 1,48773

Tests for Randomness of residuals
Box-Pierce Test
     Test based on first 12 autocorrelations
     Large sample test statistic = 9,26715
     P-value = 0,597248
```

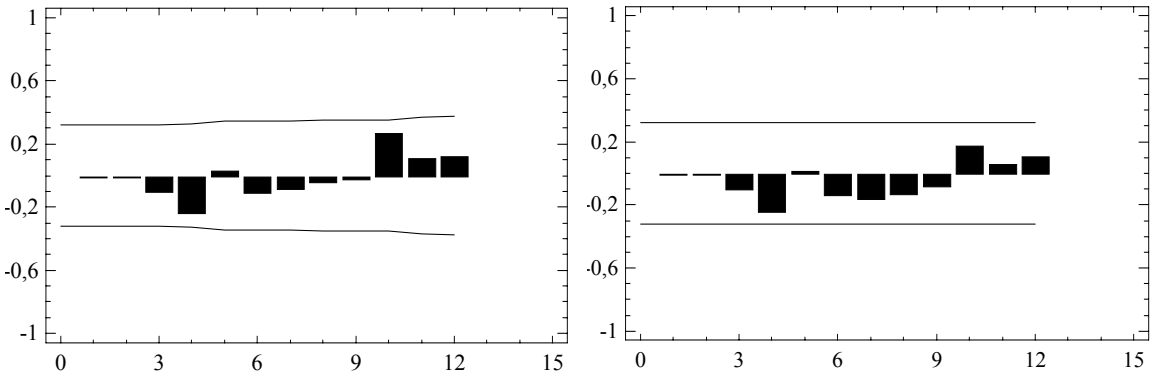


Obr. 3.11h,i: Reziduální ACF a PACF modelu ARIMA(1,1,0)$_c$

I když reziduální ACF a PACF (obr. 3.11h,i) indikují, že nesystematická složka je typu bílého šumu a rovněž hodnota "P-value" Boxova-Pierceova testu je relativně vysoká (0,597248), provedeme i odhad modelu ARIMA(0,1,1)$_c$, jeho výsledky obsahuje tab. 3.3c.

Tab. 3.3c. Výstupní tabulka modelu ARIMA(0,1,1)$_c$

```
Nonseasonal differencing of order: 1
Forecast model selected: ARIMA(0,1,1) with constant
Number of forecasts generated: 3  Number of periods withheld for validation: 0

            Estimation      Validation
Statistic   Period          Period
--------------------------------------------
MSE         2,05362
MAE         0,9854
MAPE        1,32419
ME          0,00350342
MPE         -0,0689866

                            ARIMA Model Summary
Parameter           Estimate        Stnd. Error     t               P-value
-----------------------------------------------------------------------------
MA(1)               -0,569874       0,147994        -3,85066        0,000480
Mean                1,67905         0,362276        4,63471         0,000048
Constant            1,67905
-----------------------------------------------------------------------------
Estimated white noise variance = 2,054 with 35 degrees of freedom
Estimated white noise standard deviation = 1,43318
Number of iterations: 3

Tests for Randomness of residuals
Box-Pierce Test
     Test based on first 12 autocorrelations
     Large sample test statistic = 7,06119
     P-value = 0,794106
```

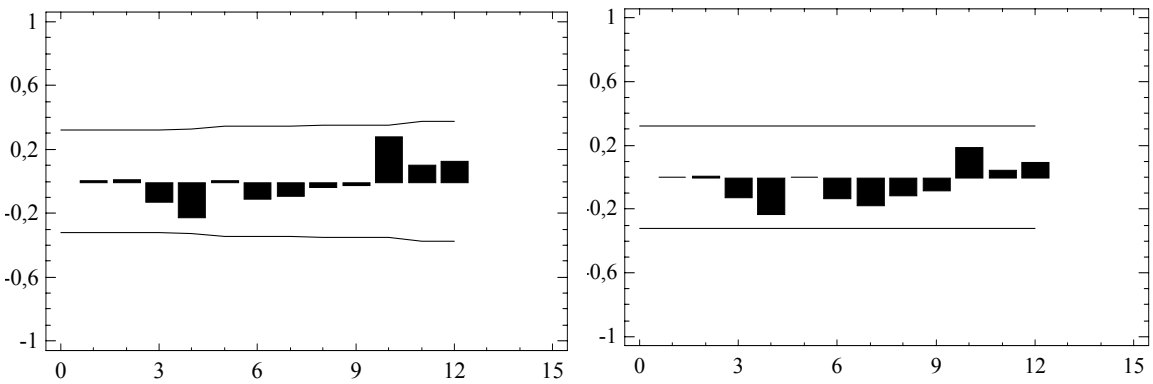


Obr. 3.11j,k: Reziduální ACF a PACF modelu ARIMA(0,1,1)$_c$

Reziduální ACF a PACF (obr. 3.11j,k) tohoto modelu také indikují, že nesystematická složka je typu bílého šumu, hodnota "P-value" Boxova-Pierceova testu (0,794106) je dokonce vyšší než u modelu s částí AR(1). Pro úplnost odhadneme ještě model obsahující obě složky, tj. AR(1) a MA(1).

Tab. 3.3d. Výstupní tabulka modelu ARIMA(1,1,1)$_c$

```
Nonseasonal differencing of order: 1
Forecast model selected: ARIMA(1,1,1) with constant
Number of forecasts generated: 3
Number of periods withheld for validation: 0

            Estimation      Validation
Statistic   Period          Period
--------------------------------------------
MSE         2,10977
MAE         0,98844
MAPE        1,32751
ME          0,00204376
MPE         -0,0731281

                              ARIMA Model Summary
Parameter           Estimate        Stnd. Error     t               P-value
--------------------------------------------------------------------------------
AR(1)               -0,0807671      0,312105        -0,258782       0,797364
MA(1)               -0,635226       0,246258        -2,57952        0,014390
Mean                1,6766          0,361225        4,64143         0,000050
Constant            1,81201
--------------------------------------------------------------------------------
Backforecasting: yes
Estimated white noise variance = 2,10992 with 34 degrees of freedom
Estimated white noise standard deviation = 1,45256
Number of iterations: 6

Tests for Randomness of residuals
Box-Pierce Test
     Test based on first 12 autocorrelations
     Large sample test statistic = 7,25441
     P-value = 0,701229
```



Obr. 3.11l,m: Reziduální ACF a PACF modelu ARIMA(1,1,1)$_c$

I když reziduální ACF a PACF, které jsou na obr. 3.11l,m indikují, že rezidua mají nesystematický charakter a hodnota "P-value" Boxova-Pierceova testu je relativně vysoká (0,701229), z *t*-testů parametrů modelu je zřejmé, že část AR(1) do modelu nepatří.

Nejen tento závěr však svědčí ve prospěch modelu ARIMA(0,1,1)$_c$. Ve srovnání s modelem ARIMA(1,1,0)$_c$ tento model vede k vyšší hodnotě "P-value" Boxova-Pierceova testu, a k nižším hodnotám MSE, MAE, MAPE, ME a MPE.

Po porovnání hodnot "P-value" parametrů $c = \mu$ (0,000048) a θ_1 (0,000480) s hladinou významnosti α (0,05) u provedeného t-testu z tab. 3.3c je zřejmé, že oba parametry jsou nenulové a model s odhadnutými parametry má tvar

$$(1 - B)y_t = 1{,}67905 + (1 - (-0{,}569874)B)a_t,$$

lze jej vyjádřit také jako

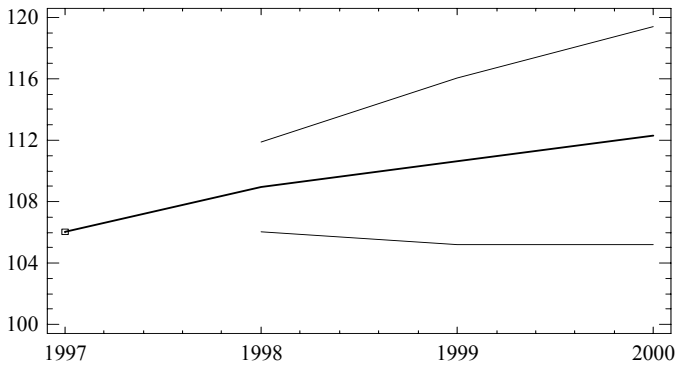
$$y_t = 1{,}67905 + y_{t-1} + a_t + 0{,}569874a_{t-1}.$$

Tab. 3.3e obsahuje bodové a intervalové předpovědi časové řady na roky 1998 - 2000.

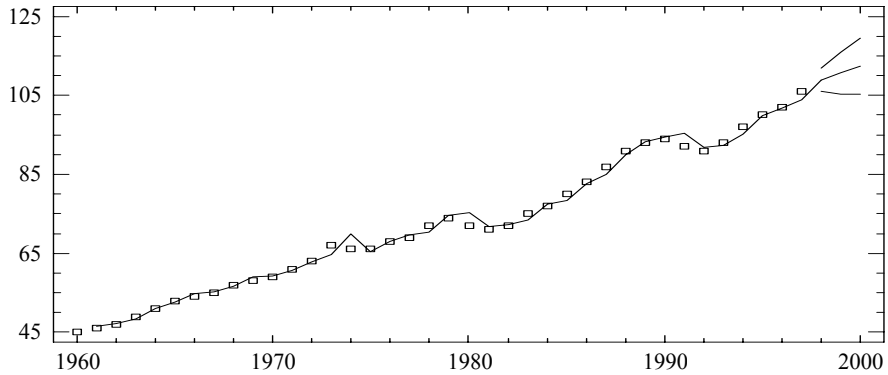
Tab. 3.3e. Předpovědi, model ARIMA(0,1,1)$_c$

```
Forecast Table for HDP_VB
Model: ARIMA(0,1,1) with constant

                                    Lower 95,0%         Upper 95,0%
Period          Forecast            Limit               Limit
-------------------------------------------------------------------
1998            108,942             106,032             111,851
1999            110,621             105,205             116,036
2000            112,3               105,215             119,385
```

Graf předpovědí je uveden na obr. 3.11n a graf časové řady s vyrovnanými hodnotami a předpověďmi je zachycen na obr. 3.11o.

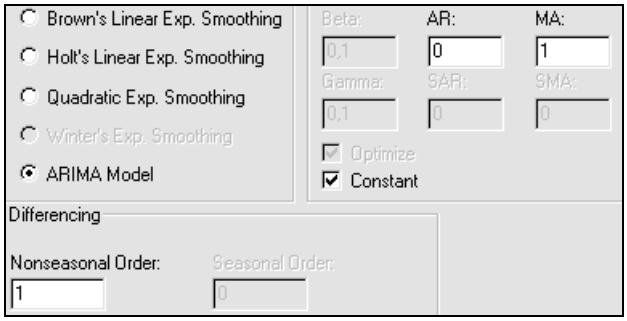


Obr. 3.11n: Předpovědi



Obr. 3.11o: Časová řada s předpověďmi

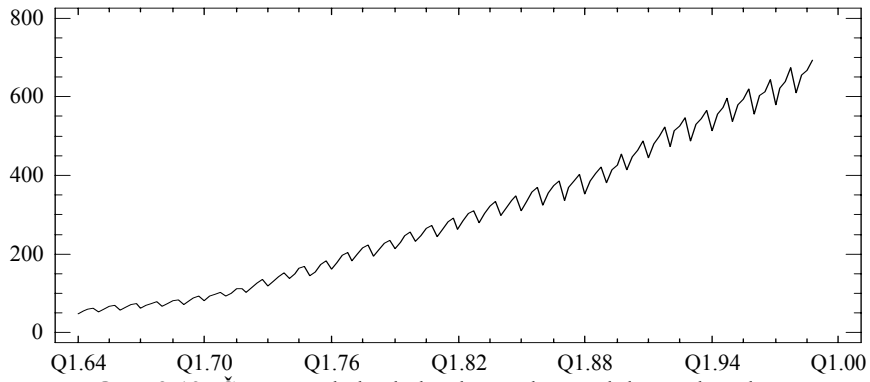
1) Vstupní panel vyplníme stejně jako v předchozím příkladu s tím rozdílem, že do řádku Data vložíme časovou řadu HDP_VB, v části **Sampling Interval** zapíšeme do políčka Starting At hodnotu 1960 a do políčka Number of Forecasts zadáme 3 předpovědi.
2) V doplňkovém panelu ***Model Specifications Options*** (obr. 3.7) v části **Type** zvolíme ARIMA Model a do všech políček v **Parameters and Terms** a **Differencing** zadáme hodnotu 0.
3) Z vybereme graf časové řady (obr. 3.11), graf autokorelační funkce a parciální autokorelační funkce (obr. 3.11a,b) a periodogram (obr. 3.11c). Z navíc vybereme testy nahodilosti reziduí.
4) Podle grafů ACF, PACF a periodogramu původní časové řady zvolíme pro transformaci časové řady nesezónní diferenci přepsáním nuly na jedničku v políčku Nonseasonal Order v části **Differencing**. Odhadnutý model ARIMA$(0,1,0)_c$ (tab. 3.3a) se objeví automaticky ve výsledkovém okně, současně se přepočítá Boxův-Pierceův test a podle vypočítaných reziduí se překreslí všechny grafy (obr. 3.11d-g).
5) Z grafů reziduální ACF a PACF modelu ARIMA$(0,1,0)_c$ je patrné, že bude třeba model dále rozšířit, není však zcela zřejmé jestli o část AR(1) nebo MA(1).
   a) Přidáme tedy do modelu nejprve část AR(1), tak že v doplňkovém panelu ***Model Specifications Options*** přepíšeme v políčku AR v **Parameters and Terms** nulu na jedničku, a odhadneme model ARIMA$(1,1,0)_c$ (tab. 3.3b) a zkontrolujeme překreslené grafy (obr. 3.11h,i).
   b) V doplňkovém panelu zrušíme nastavení AR(1) a přidáme do modelu část MA(1), přepsáním nuly na jedničku v políčku MA, odhadneme tak model ARIMA$(0,1,1)_c$ (tab. 3.3c) a zkontrolujeme překreslené grafy (obr. 3.11j,k).
   c) Pro úplnost odhadneme model obsahující současně AR(1) i MA(1), tj. model ARIMA$(1,1,1)_c$. V doplňkovém panelu zapíšeme jedničku do políček AR i MA a získáme model z tab. 3.3d.
6) V předchozím kroku jsme na základě hodnoty Boxova-Pierceova testu a hodnot MSE, MAE, MAPE, ME a MPE vybrali jako nejlepší model ARIMA$(0,1,1)_c$. Vypočítáme předpovědi (tab. 3.3e) a zobrazíme graf předpovědí (obr. 3.11n) a graf časové řady s vyrovnanými hodnotami a předpověďmi (obr. 3.11o).



Obr. 3.11p: Výřez vyplněného doplňkového panelu Model Specifications Options pro model ARIMA$(0,1,1)_c$

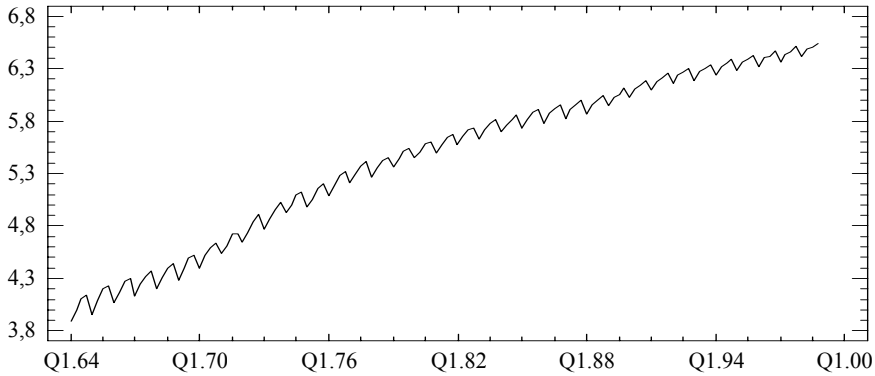
Příklad 3.3

Pomocí Boxovy-Jenkinsovy metodologie najděte model čtvrtletní časové řady hrubého domácího produktu Rakouska (HDP_RAK), která je k dispozici v mld. ATS od I. čtvrtletí roku 1964 do IV. čtvrtletí roku 1998. Na základě zvoleného modelu vypočítejte předpovědi této časové řady na 4 roky.

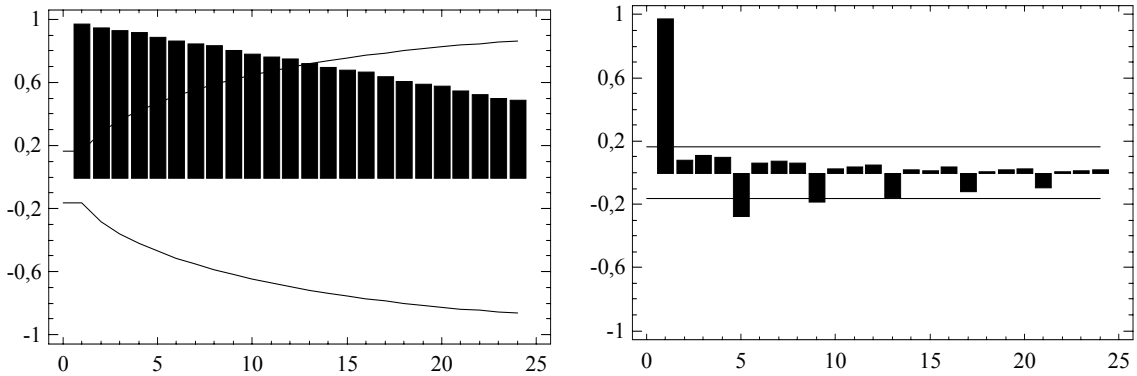


Obr. 3.12: Časová řada hrubého domácího produktu Rakouska

Průběh časové řady je zachycen na obr. 3.12. Časová řada je nestacionární a obsahuje sezónní složku. Protože tato řada vykazuje mírně exponenciální průběh s proporcionální sezónností, je vhodné ji logaritmovat. Takto transformovaná časová řada je na obr. 3.12a.

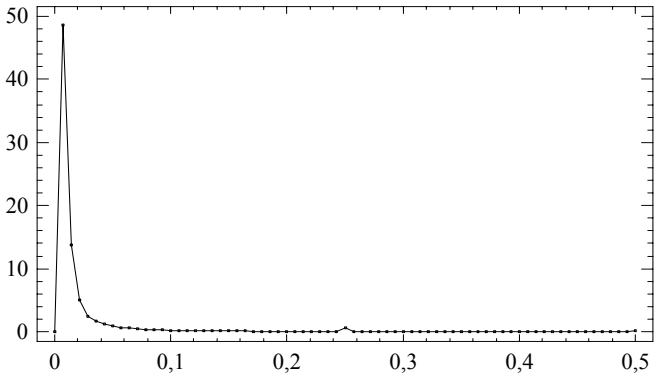


Obr. 3.12a: Logaritmovaná časová řada hrubého domácího produktu Rakouska



Obr. 3.12b,c: ACF a PACF logaritmované časové řady

Logaritmovaná časová řada je zjevně nestacionární, což potvrzuje i tvar ACF a PACF, který je na obr. 3.12b,c. Hodnoty ACF klesají velmi pomalu a první hodnota, stejně jako u PACF, je blízká jedné. Periodogram, zachycený na obr. 3.12d má významný vrchol v nulové (nesezónní) frekvenci.



Obr. 3.12d: Periodogram logaritmované časové řady

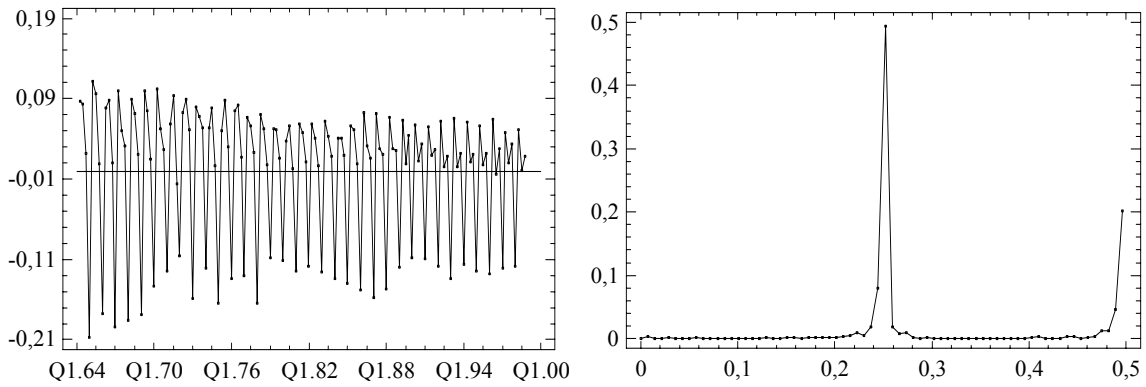
I když je z grafu logaritmované časové řady a tvaru PACF zřejmé, že časová řada obsahuje sezónní složku, ACF ani periodogram její přítomnost na první pohled nepotvrzují. Identifikační hodnoty sezónnosti u ACF splývají s identifikačními hodnotami pro stochastický trend a sezónnost se na tvaru ACF projevuje pouze tím, že klesání jejích hodnot není plynulé, ale nepatrně schodovité. Dominantnost stochastického trendu v časové řadě je zřetelná i na tvaru periodogramu, kde téměř zcela zakrývá identifikační hodnoty sezónnosti ve frekvencích π a $\pi/2$.

Z uvedeného je zřejné, že časovou řadu budeme stacionarizovat I. nesezónní diferencí. Odhad modelu ARIMA$(0,1,0)_c$ obsahuje tab. 3.4a.

Tab. 3.4a. Výstupní tabulka modelu ARIMA$(0,1,0)_c$

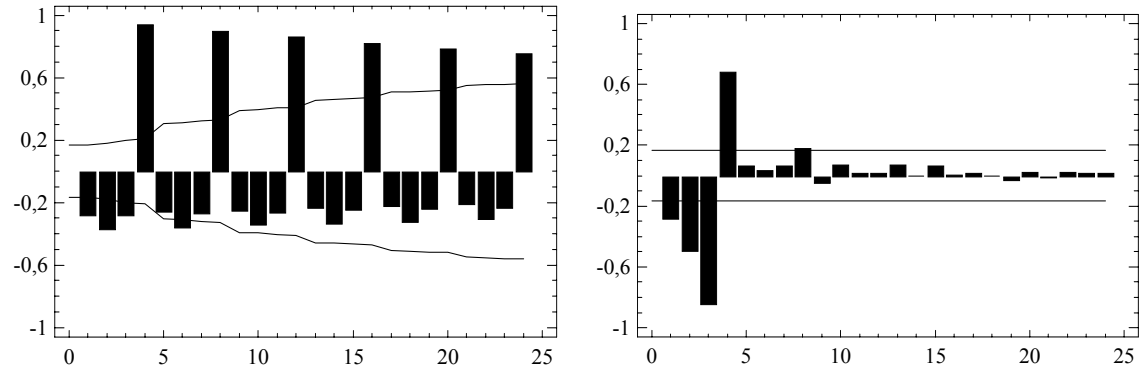
```
Math adjustment: Natural log
Nonseasonal differencing of order: 1
Forecast model selected: ARIMA(0,1,0) with constant
Number of forecasts generated: 12  Number of periods withheld for validation: 0

            Estimation      Validation
Statistic   Period          Period
--------------------------------------------
MSE         617,148
MAE         17,6602
MAPE        6,93991
ME          -1,02455
MPE         -0,358948

                            ARIMA Model Summary
Parameter           Estimate        Stnd. Error     t               P-value
-----------------------------------------------------------------------------
Mean                0,0190508       0,00711191      2,67871         0,008289
Constant            0,0190508
-----------------------------------------------------------------------------
Estimated white noise variance = 0,00703052 with 138 degrees of freedom
Estimated white noise standard deviation = 0,0838482
Number of iterations: 1

Tests for Randomness of residuals
Box-Pierce Test
Test based on first 24 autocorrelations
Large sample test statistic = 800,109
P-value = 0,0
```

Rezidua a reziduální periodogram časové řady po I. nesezónní diferenci zobrazuje obr. 3.12e,f. Z grafu reziduí je patrné, že po odstranění trendu v časové řadě zůstala zachována výrazná sezónnost. Totéž potvrzuje i periodogram, který má významné vrcholy v obou sezónních frekvencích π a $\pi/2$.



Obr. 3.12e,f: Rezidua a reziduální periodogram modelu ARIMA$(0,1,0)_c$

Na obr. 3.12g,h je vidět, že ACF má významnou každou čtvrtou kladnou hodnotu. Tyto hodnoty jsou blízké jedné a pomalu klesají, záporné hodnoty jsou výrazně nižší a klesají ještě pomaleji. PACF má tři významné záporné hodnoty (třetí je blízká -1) a dvě kladné.



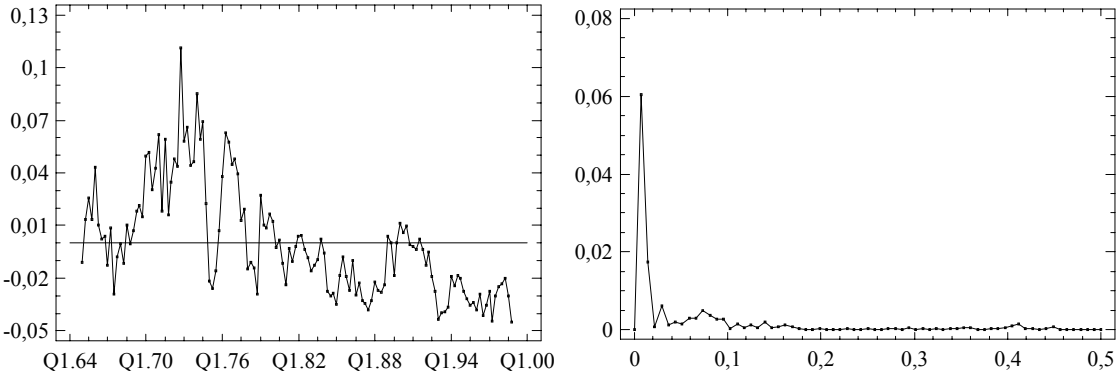
Obr. 3.12g,h: Reziduální ACF a PACF modelu ARIMA$(0,1,0)_c$

Z uvedeného lze předpokládat, že se jedná o sezónní integrovaný proces SI(1). Časovou řadu proto nebudeme stacionarizovat nesezónní diferencí, ale použijeme diferenci sezónní (sezónní diference obsahuje také nesezónní diferenci). Odhad modelu SARIMA$(0,0,0)(0,1,0)_c$ je v tab. 3.4b.

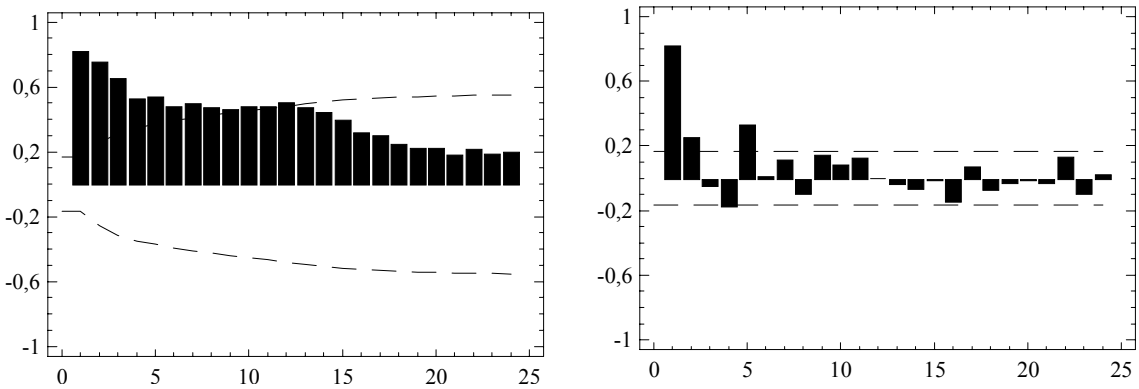
Tab. 3.4b. Výstupní tabulka modelu SARIMA(0,0,0)(0,1,0)$_c$

```
Math adjustment: Natural log
Seasonal differencing of order: 1
Forecast model selected: ARIMA(0,0,0)x(0,1,0)4 with constant
Number of forecasts generated: 12  Number of periods withheld for validation: 0

            Estimation      Validation
Statistic   Period          Period
-------------------------------------------
MSE         100,48
MAE         7,20392
MAPE        2,4285
ME          -3,77898
MPE         -0,0470433

                            ARIMA Model Summary
Parameter           Estimate        Stnd. Error     t               P-value
-----------------------------------------------------------------------------
Mean                0,0721088       0,00265227      27,1876         0,000000
Constant            0,0721088
-----------------------------------------------------------------------------
Estimated white noise variance = 0,000956698 with 135 degrees of freedom
Estimated white noise standard deviation = 0,0309305
Number of iterations: 1

Tests for Randomness of residuals
Box-Pierce Test
Test based on first 24 autocorrelations
Large sample test statistic = 669,619
P-value = 0,0
```

Rezidua a reziduální periodogram časové řady po sezónní diferenci zachycuje obr. 3.12i,j. Z grafu reziduí je zřejmé, že časová řada po sezónní diferenci není stacionární, což potvrzuje i významný vrchol v nesezónní frekvenci periodogramu.



Obr. 3.12i,j: Rezidua a reziduální periodogram modelu SARIMA(0,0,0)(0,1,0)$_c$



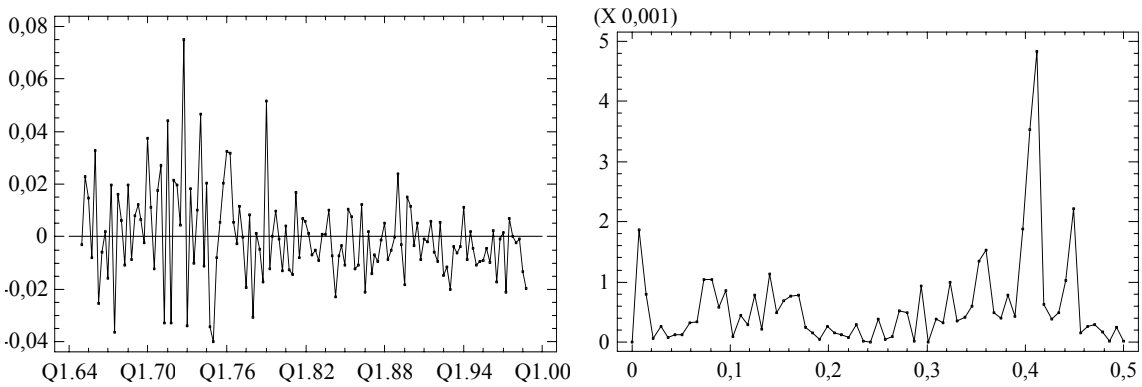
Obr. 3.12k,l: Reziduální ACF a PACF modelu SARIMA(0,0,0)(0,1,0)$_c$

Reziduální ACF a PACF je zachycena na obr. 3.12k,l. Protože hodnoty reziduální ACF vykazují na začátku spíše exponenciální pokles a PACF má statisticky významnou první hodnotu, dáme při rozšiřování modelu přednost části AR(1) před nesezónní diferencí. Budeme tak odhadovat model SARIMA(1,0,0)(0,1,0)$_c$. Výsledky jsou uvedeny v tab. 3.4c.

Tab. 3.4c. Výstupní tabulka modelu SARIMA(1,0,0)(0,1,0)$_c$

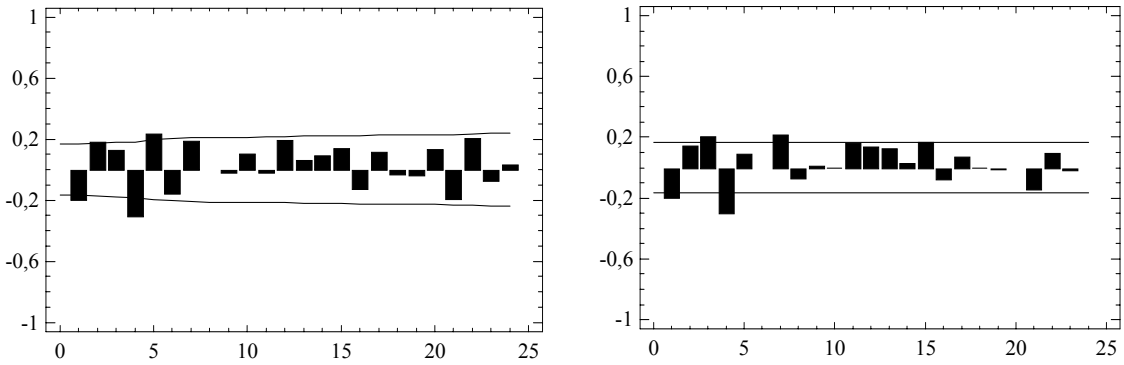
```
Math adjustment: Natural log
Seasonal differencing of order: 1
Forecast model selected: ARIMA(1,0,0)x(0,1,0)4 with constant
Number of forecasts generated: 12
Number of periods withheld for validation: 0

            Estimation      Validation
Statistic   Period          Period
--------------------------------------------
MSE         16,9669
MAE         3,05048
MAPE        1,27419
ME          -0,776394
MPE         -0,0202029

                            ARIMA Model Summary
Parameter           Estimate        Stnd. Error     t               P-value
--------------------------------------------------------------------------------
AR(1)               0,827347        0,0490933       16,8525         0,000000
Mean                0,0711517       0,00813941      8,74163         0,000000
Constant            0,0122846
--------------------------------------------------------------------------------
Backforecasting: yes
Estimated white noise variance = 0,000307532 with 134 degrees of freedom
Estimated white noise standard deviation = 0,0175366
Number of iterations: 1

Tests for Randomness of residuals
Box-Pierce Test
     Test based on first 24 autocorrelations
     Large sample test statistic = 70,3915
     P-value = 0,00000105795
```

Rezidua a reziduální periodogram modelu SARIMA(1,0,0)(0,1,0)$_c$ ukazuje obr. 3.12m,n. Oba grafy potvrzují, že rezidua jsou stacionární.



Obr. 3.12m,n: Rezidua a reziduální periodogram modelu SARIMA(1,0,0)(0,1,0)$_c$

Grafy reziduálních ACF a PACF, které jsou na obr. 3.12o,p neindikují nesystematičnost (první hodnoty a některé další leží vně tolerančních mezí).



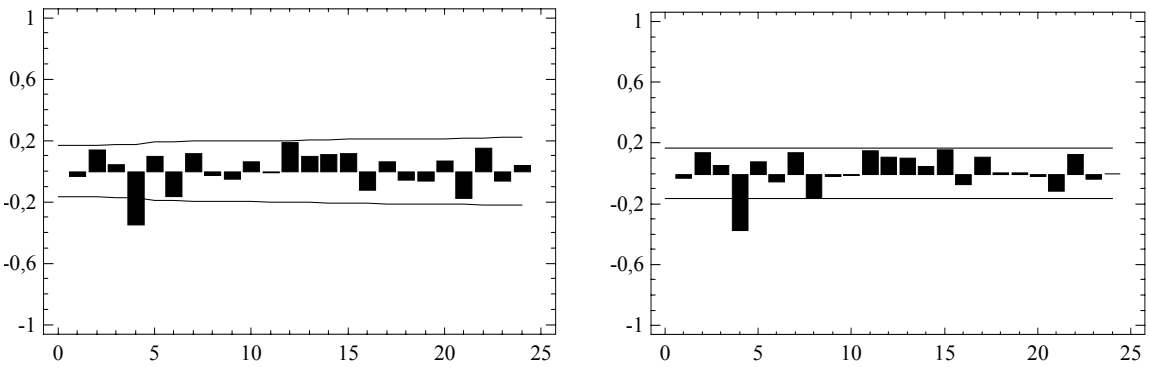
Obr. 3.12o,p: Reziduální ACF a PACF modelu SARIMA(1,0,0)(0,1,0)$_c$

Zkusíme tedy doplnit model ještě o část MA(1). Odhadnutý model SARIMA(1,0,1)(0,1,0)$_c$ ukazuje tab. 3.4d.

Tab. 3.4d. Výstupní tabulka modelu SARIMA(1,0,1)(0,1,0)$_c$

```
Math adjustment: Natural log
Seasonal differencing of order: 1
Forecast model selected: ARIMA(1,0,1)x(0,1,0)4 with constant
Number of forecasts generated: 12  Number of periods withheld for validation: 0

            Estimation      Validation
Statistic   Period          Period
--------------------------------------------
MSE         16,0048
MAE         2,92953
MAPE        1,22599
ME          -0,597773
MPE         -0,0139752

                            ARIMA Model Summary
Parameter           Estimate        Stnd. Error     t               P-value
-----------------------------------------------------------------------------
AR(1)               0,905701        0,0440679       20,5524         0,000000
MA(1)               0,237728        0,0983739       2,41657         0,017024
Mean                0,069582        0,0118679       5,86305         0,000000
Constant            0,00656153
-----------------------------------------------------------------------------
Estimated white noise variance = 0,000294075 with 133 degrees of freedom
Estimated white noise standard deviation = 0,0171486

Tests for Randomness of residuals
Box-Pierce Test
     Test based on first 24 autocorrelations
     Large sample test statistic = 49,6424
     P-value = 0,000655342
```

Rezidua a reziduální periodogram modelu SARIMA(1,0,1)(0,1,0)$_c$ jsou obdobná jako u modelu SARIMA(1,0,0)(0,1,0)$_c$.



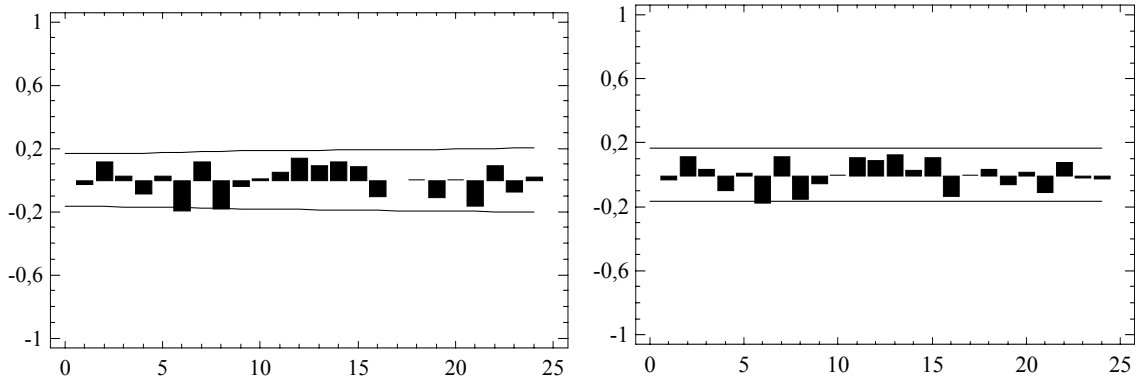
Obr. 3.12q,r: Reziduální ACF a PACF modelu SARIMA(1,0,1)(0,1,0)$_c$

I když se hodnota Boxova-Pierceova testu zvýšila (0,000655342), grafy reziduální ACF a PACF na obr. 3.12q,r stále potvrzují existenci systematické složky. U obou funkcí je patrná statisticky významná čtvrtá hodnota, pokusíme se proto dále rozšířit model o SAR(1) a odhadneme tak model SARIMA(1,0,1)(1,1,0)$_c$. Výsledky jsou uvedeny v tab. 3.4e.

Tab. 3.4e. Výstupní tabulka modelu SARIMA(1,0,1)(1,1,0)$_c$

```
Math adjustment: Natural log
Seasonal differencing of order: 1
Forecast model selected: ARIMA(1,0,1)x(1,1,0)4 with constant
Number of forecasts generated: 12
Number of periods withheld for validation: 0

            Estimation      Validation
Statistic   Period          Period
--------------------------------------------
MSE         14,715
MAE         2,80309
MAPE        1,15331
ME          -0,467416
MPE         -0,021032

                            ARIMA Model Summary
Parameter           Estimate        Stnd. Error     t               P-value
--------------------------------------------------------------------------
AR(1)               0,96044         0,0283191       33,9149         0,000000
MA(1)               0,265289        0,089951        2,94927         0,003769
SAR(1)              -0,415108       0,0805285       -5,1548         0,000001
Mean                0,0665332       0,017835        3,73048         0,000283
Constant            0,00372459
--------------------------------------------------------------------------
Backforecasting: yes
Estimated white noise variance = 0,000251297 with 132 degrees of freedom
Estimated white noise standard deviation = 0,0158523
Number of iterations: 6

Tests for Randomness of residuals
Box-Pierce Test
     Test based on first 24 autocorrelations
     Large sample test statistic = 30,4299
     P-value = 0,0836842
```

Po doplnění modelu o část SMA(1) se hodnoty reziduální ACF a PACF, které jsou na obr. 3.12s,t snížily tak, že všechny leží uvnitř nebo na hranici intervalu spolehlivosti. Toto potvrzuje i hodnota Boxova-Pierceova testu, která se oproti minulému modelu zvýšila (0,0836842).



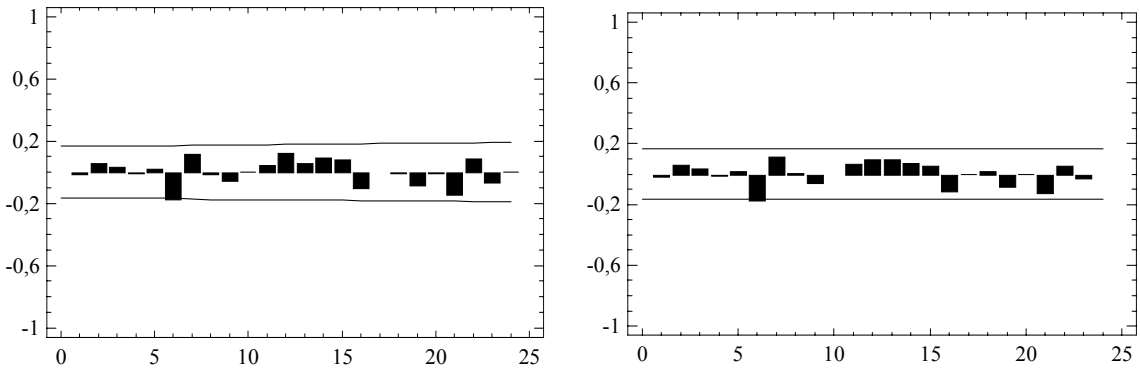
Obr. 3.12s,t: Reziduální ACF a PACF modelu SARIMA(1,0,1)(1,1,0)$_c$

Nyní se pokusíme model SARIMA(1,0,1)(0,1,0)$_c$ rozšířit místo o část SAR(1) o část SMA(1). Odhadnutý model SARIMA(1,0,1)(0,1,1)$_c$ ukazuje tab. 3.4f.

Tab. 3.4f. Výstupní tabulka modelu SARIMA(1,0,1)(0,1,1)$_c$

```
Math adjustment: Natural log
Seasonal differencing of order: 1
Forecast model selected: ARIMA(1,0,1)x(0,1,1)4 with constant
Number of forecasts generated: 12  Number of periods withheld for validation: 0

            Estimation      Validation
Statistic   Period          Period
-------------------------------------------
MSE         14,0419
MAE         2,74708
MAPE        1,12502
ME          -0,499757
MPE         -0,0413263

                            ARIMA Model Summary
Parameter           Estimate        Stnd. Error     t               P-value
-----------------------------------------------------------------------------
AR(1)               0,982932        0,0207264       47,4242         0,000000
MA(1)               0,277653        0,0862714       3,21836         0,001622
SMA(1)              0,518678        0,0770593       6,73089         0,000000
Mean                0,0631601       0,0259651       2,4325          0,016335
Constant            0,00107799
-----------------------------------------------------------------------------
Estimated white noise variance = 0,000238628 with 132 degrees of freedom
Estimated white noise standard deviation = 0,0154476

Tests for Randomness of residuals
Box-Pierce Test
     Test based on first 24 autocorrelations
     Large sample test statistic = 19,1568
     P-value = 0,575081
```

Hodnoty reziduální ACF a PACF se ve srovnání s předchozím modelem ještě snížily a lze předpokládat, že vykazují nesystematický pohyb. Toto potvrzuje i hodnota Boxova-Pierceova testu, která se oproti minulému modelu výrazně zvýšila (0,575081).



Obr. 3.12u,v: Reziduální ACF a PACF modelu SARIMA(1,0,1)(0,1,1)$_c$

Porovnáním hodnot "P-value" parametrů ϕ_1 (0,000000), θ_1 (0,001622), μ (0,016335) a Θ_1 (0,000000) s hladinou významnosti α (0,05) u provedeného t-testu v tab. 3.4f, prokážeme hypotézu, že všechny parametry včetně konstanty $c = \mu(1 - \phi_1 - \phi_2)$ jsou nenulové a odhadnutý model má formu

$$(1 - 0{,}982932B)(1 - B^4)y_t = 0{,}00107799 + (1 - 0{,}518678B^4)(1 - 0{,}277653B)a_t,$$

lze jej vyjádřit také jako

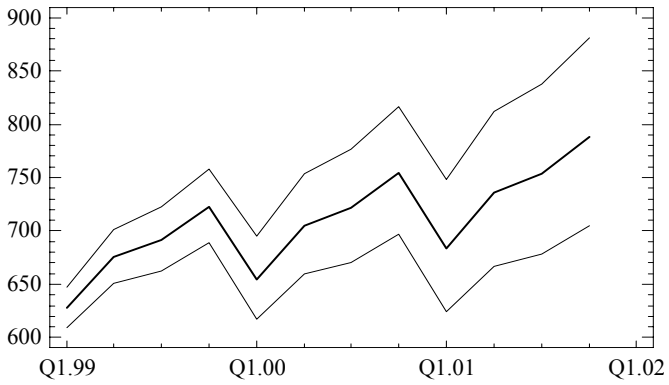
$$y_t = 0{,}00107799 + 0{,}982932y_{t-1} + y_{t-4} - 0{,}982932y_{t-5} + \\ + a_t - 0{,}277653a_{t-1} - 0{,}518678a_{t-4} + 0{,}1440125a_{t-5}.$$

Tab. 3.4g obsahuje bodové a intervalové předpovědi časové řady na jednotlivá čtvrtletí let 1999 - 2001.

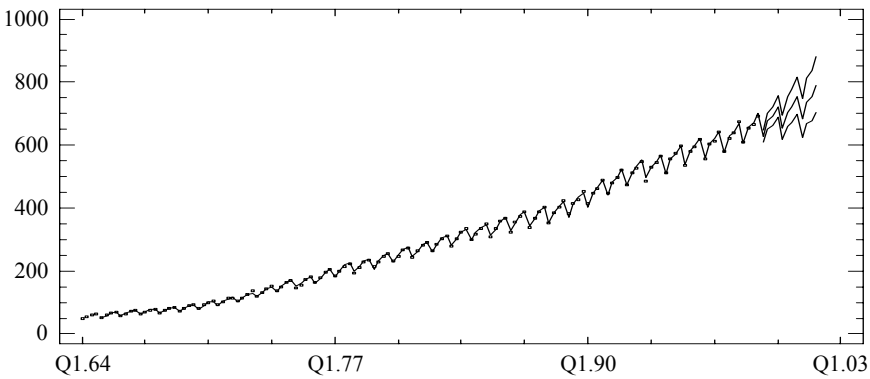
Tab. 3.4g. Předpovědi

```
Forecast Table for HDP_RAK
Model: ARIMA(1,0,1)x(0,1,1)4 with constant
       Math adjustment: Natural log

                                    Lower 95,0%         Upper 95,0%
Period          Forecast            Limit               Limit
-------------------------------------------------------------------------------
Q1.99           628,156             609,252             647,647
Q2.99           675,821             651,017             701,569
Q3.99           691,742             662,644             722,118
Q4.99           722,689             688,989             758,038
Q1.00           654,795             617,089             694,806
Q2.00           704,741             659,276             753,341
Q3.00           721,606             670,729             776,341
Q4.00           754,158             696,994             816,009
Q1.01           683,547             624,646             748,002
Q2.01           735,939             667,157             811,812
Q3.01           753,806             678,476             837,499
Q4.01           788,072             704,733             881,267
```

Graf předpovědí je uveden na obr. 3.12w a graf časové řady s vyrovnanými hodnotami a předpověďmi je zachycen na obr. 3.12x.

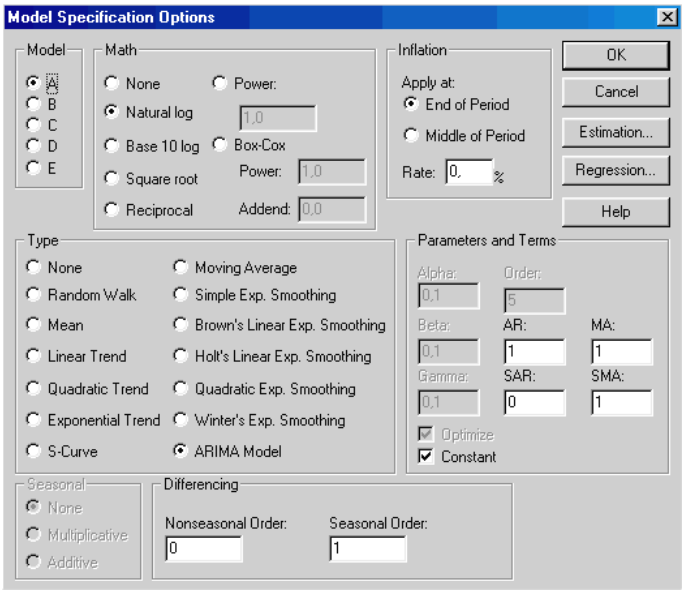


Obr. 3.12w: Předpovědi



Obr. 3.12x: Časová řada s předpověďmi

1) Ve vstupním panelu vložíme do řádku Data časovou řadu HDP_RAK a v části **Sampling Interval** vybereme Quarter(s). V políčku Starting At přepíšeme nabízenou hodnotu na Q1.64 a do políčka Seasonality zadáme délku sezóny 4. Počet předpovědí 12 zadáme do políčka Number of Forecasts.
2) Pro provedení logaritmické transformace vybereme v doplňkovém panelu ***Model Specifications Options*** (obr. 3.7) v části **Math** položku Natural log.
3) Ve stejném panelu, ale v části **Type** zvolíme ARIMA Model a do všech políček v **Parameters and Terms** a **Differencing** zadáme hodnotu 0.
4) Z vybereme graf časové řady (obr. 3.12a), graf reziduí, graf autokorelační funkce a parciální autokorelační funkce (obr. 3.12b,c) a periodogram (obr. 3.12d) a z vybereme navíc testy nahodilosti reziduí.
5) Podle grafů ACF, PACF a periodogramu logaritmované časové řady zvolíme pro transformaci nesezónní diferenci přepsáním nuly na jedničku v políčku Nonseasonal Order v části **Differencing**. Odhadnutý model ARIMA$(0,1,0)_c$ (tab. 3.4a) se objeví automaticky ve výsledkovém okně, současně se přepočítá Boxův-Pierceův test a podle vypočítaných reziduí se překreslí všechny grafy (obr. 3.12e-h).
6) Na základě grafů reziduální ACF a PACF a reziduálního periodogramu modelu ARIMA$(0,1,0)_c$ dáme před nesezónní diferencí přednost diferenci sezónní. Změníme tedy hodnotu v políčku Nonseasonal Order opět na nulu a na jedničku přepíšeme nulu v políčku Seasonal Order. Odhadneme model SARIMA$(0,0,0)(0,1,0)_c$ (tab. 3.4b) a zkontrolujeme překreslené grafy (obr. 3.12i-l).
7) Z grafů reziduální ACF a PACF modelu SARIMA$(0,0,0)(0,1,0)_c$ je patrné, že bude vhodné rozšířit model o část AR(1). Přepíšeme proto v políčku AR nulu na jedničku. Odhadneme model SARIMA$(1,0,0)(0,1,0)_c$ (tab. 3.4c) a zkontrolujeme překreslené grafy (obr. 3.12m-p).
8) Grafy reziduální ACF a PACF modelu SARIMA$(1,0,0)(0,1,0)_c$ indikují, že bude vhodné přidat do modelu část MA(1). Přepíšeme tedy v políčku MA nulu na jedničku. Odhadneme model SARIMA$(1,0,1)(0,1,0)_c$ (tab. 3.4d) a zkontrolujeme překreslené grafy (obr. 3.12q,r).
9) Podle grafů reziduální ACF a PACF modelu SARIMA$(1,0,1)(0,1,0)_c$ rozšíříme model o sezónní část SAR(1). V políčku SAR přepíšeme nulu na jedničku a odhadneme model SARIMA$(1,0,1)(1,1,0)_c$ (tab. 3.4e) a zkontrolujeme překreslené grafy (obr. 3.12s,t).
10) Vložíme do modelu místo části SAR(1) část SMA(1). V políčku SAR přepíšeme zpět jedničku na nulu a naopak v políčku SMA nulu na jedničku a odhadneme model SARIMA$(1,0,1)(0,1,1)_c$ (tab. 3.4f) a zkontrolujeme překreslené grafy (obr. 3.12u,v).



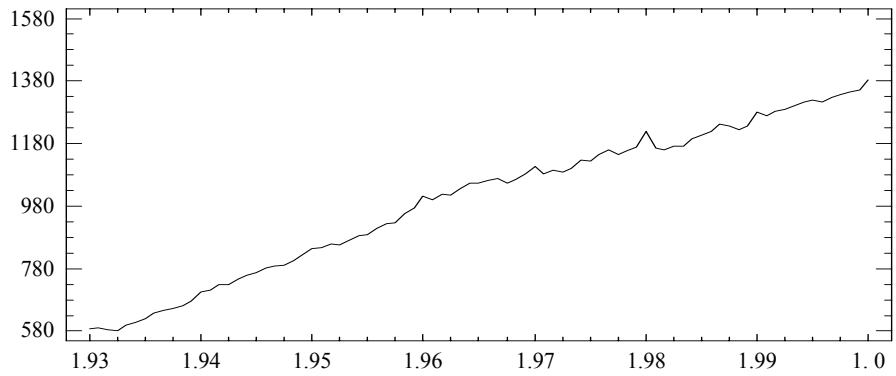
Obr. 3.12y: Doplňkový panel Model Specifications Options pro model SARIMA(1,0,1)(0,1,1)$_c$

11) Na základě zvoleného modelu SARIMA(1,0,1)(0,1,1)$_c$ vypočítáme předpovědi (tab. 3.4g). Zobrazíme graf předpovědí (obr. 3.12w) a graf časové řady s vyrovnanými hodnotami a předpověďmi (obr. 3.12x).

Příklad 3.4

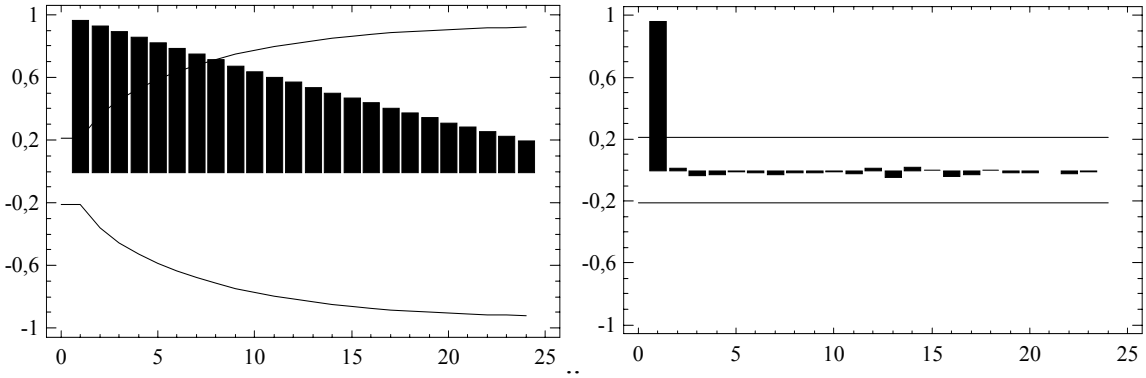
Pro měsíční časovou řadu M2 České republiky (M2_CR), která je k dispozici v mld. Kč od ledna roku 1993 do ledna roku 2000 najděte vhodný model Boxovy-Jenkinsovy metodologie a na jeho základě vypočítejte předpovědi této časové řady na 1 rok.

Průběh časové řady je zachycen na obr. 3.13. Z grafu je patrné, že časová řada je nestacionární, ale není zcela zřejmé, zda obsahuje sezónní složku.

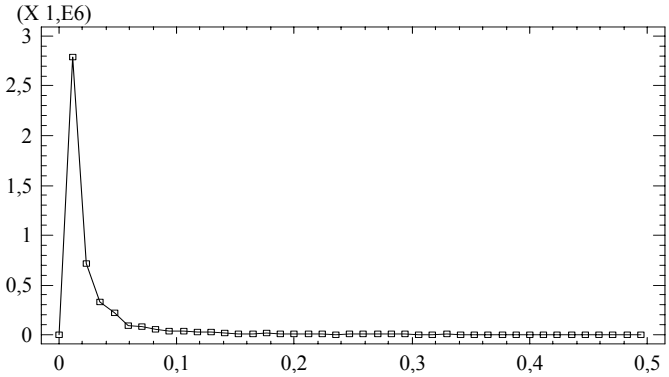


Obr. 3.13: Časová řada M2 České republiky

Nestacionaritu časové řady potvrzuje i tvar ACF a PACF. Hodnoty ACF klesají velmi pomalu a první hodnota, stejně jako u PACF, je blízká jedné. Periodogram má významný vrchol v nulové (nesezónní) frekvenci. Sezónnost neindikuje ani ACF a PACF, ani periodogram.



Obr. 3.13a,b: ACF a PACF původní časové řady



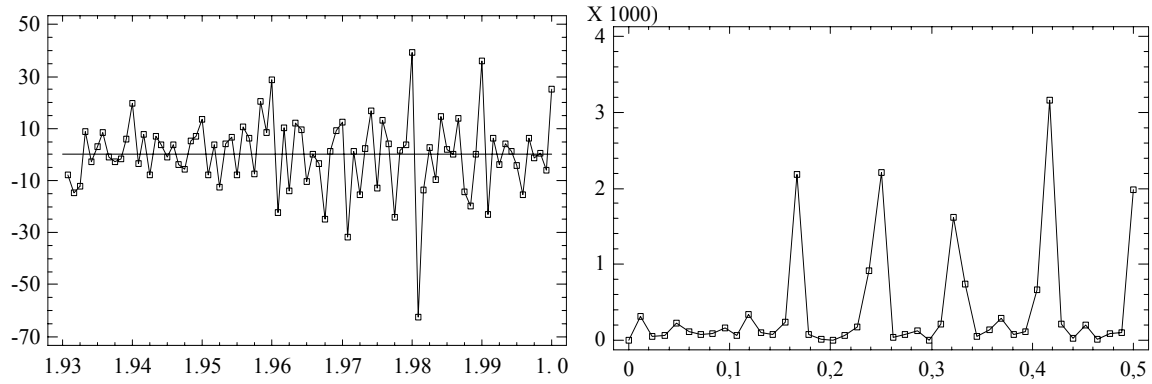
Obr. 3.13c: Periodogram původní časové řady

Časovou řadu budeme tedy stacionarizovat I. nesezónní diferencí. Odhad modelu ARIMA(0,1,0)$_c$ obsahuje tab. 3.5a.

Tab. 3.5a. Výstupní tabulka modelu ARIMA(0,1,0)$_c$

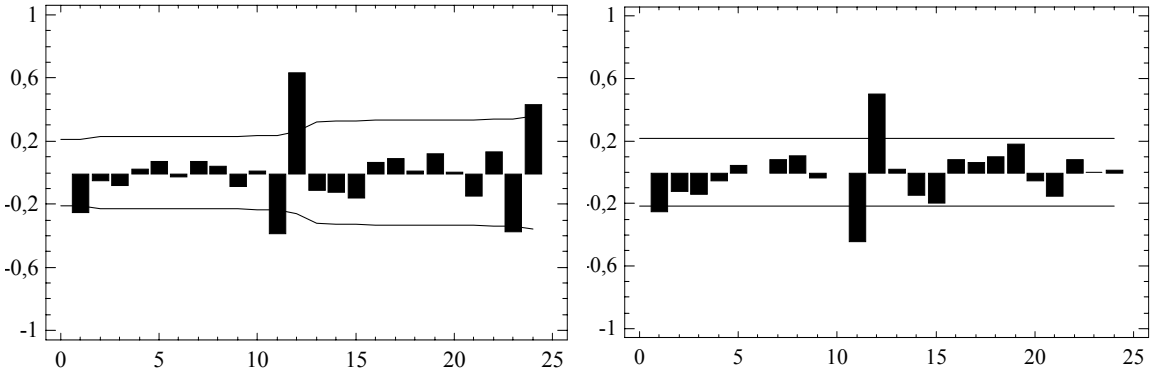
```
Nonseasonal differencing of order: 1
Forecast model selected: ARIMA(0,1,0) with constant
Number of forecasts generated: 12
Number of periods withheld for validation: 0

            Estimation      Validation
Statistic   Period          Period
--------------------------------------------
MSE         211,879
MAE         10,3401
MAPE        1,0354
ME          1,89478E-14
MPE         -0,000218943

                            ARIMA Model Summary
Parameter           Estimate        Stnd. Error     t               P-value
--------------------------------------------------------------------------------
Mean                9,48571         1,5882          5,97263         0,000000
Constant            9,48571
--------------------------------------------------------------------------------
Backforecasting: yes
Estimated white noise variance = 211,879 with 83 degrees of freedom
Estimated white noise standard deviation = 14,5561
Number of iterations: 1

Tests for Randomness of residuals
Box-Pierce Test
     Test based on first 24 autocorrelations
     Large sample test statistic = 91,4656
     P-value = 8,23501E-10
```

Rezidua a reziduální periodogram časové řady po nesezónní diferenci zobrazuje obr. 3.13d,e. Z grafu reziduí je patrné, že časová řada je po nesezónní diferenci stacionární. Totéž potvrzuje i periodogram, který navíc odhaluje, že časová řada obsahuje sezónní složku.



Obr. 3.13d,e: Rezidua a reziduální periodogram modelu ARIMA(0,1,0)$_c$

Přítomnost sezónní složky potvrzují i reziduální ACF a PACF (obr. 3.13f,g), obě funkce mají statisticky významnou každou jedenáctou zápornou hodnotu a každou dvanáctou kladnou hodnotu, současně však je významná i první hodnota.



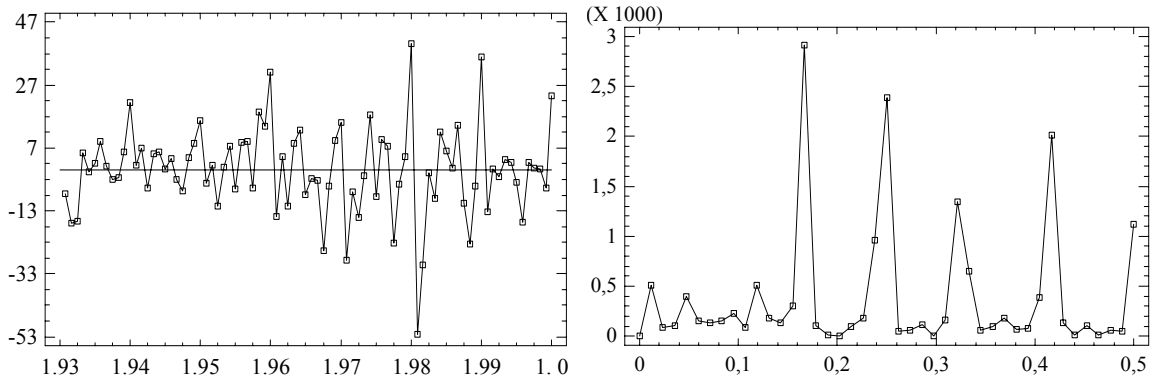
Obr. 3.13f,g: Reziduální ACF a PACF modelu ARIMA(0,1,0)$_c$

Z toho důvodu rozšíříme model nejprve o část AR(1). Budeme odhadovat model ARIMA(1,1,0)$_c$.

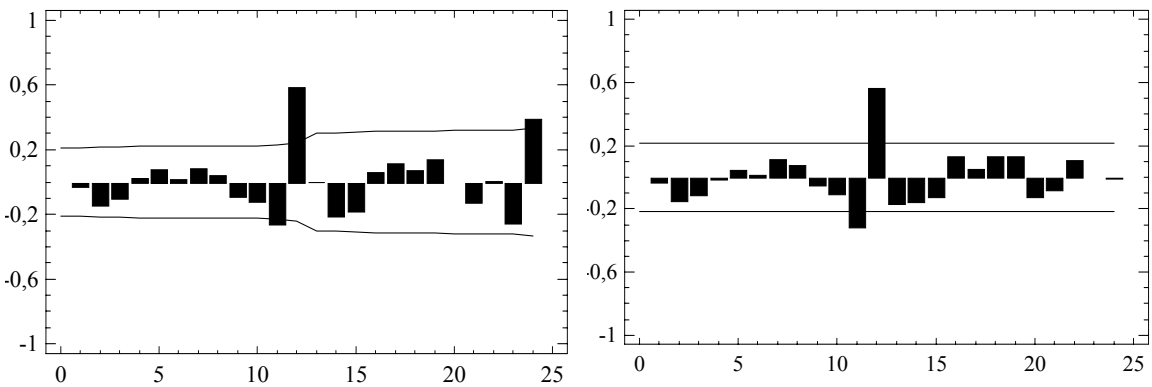
Tab. 3.5b. Výstupní tabulka modelu ARIMA(1,1,0)$_c$

```
Nonseasonal differencing of order: 1
Forecast model selected: ARIMA(1,1,0) with constant
Number of forecasts generated: 12
Number of periods withheld for validation: 0

            Estimation      Validation
Statistic   Period          Period
-------------------------------------------
MSE         199,781
MAE         10,0064
MAPE        1,00242
ME          -0,0358967
MPE         0,0000181012

                              ARIMA Model Summary
Parameter           Estimate        Stnd. Error     t               P-value
-----------------------------------------------------------------------------
AR(1)               -0,263041       0,108397        -2,42665        0,017431
Mean                9,45699         1,23125         7,68079         0,000000
Constant            11,9446
-----------------------------------------------------------------------------
Backforecasting: yes
Estimated white noise variance = 199,827 with 82 degrees of freedom
Estimated white noise standard deviation = 14,136
Number of iterations: 1

Tests for Randomness of residuals
Box-Pierce Test
     Test based on first 24 autocorrelations
     Large sample test statistic = 71,4552
     P-value = 7,23453E-7
```

Rezidua a reziduální periodogram modelu ARIMA(1,1,0)$_c$ zobrazuje obr. 3.13h,i.



Obr. 3.13h,i: Rezidua a reziduální periodogram modelu ARIMA(1,1,0)$_c$



Obr. 3.13j,k: Reziduální ACF a PACF modelu ARIMA(1,1,0)$_c$

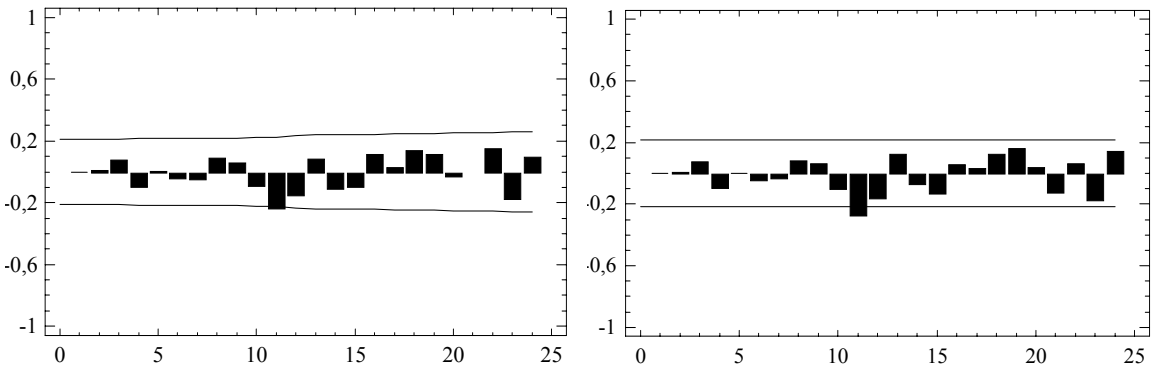
Z grafů reziduální ACF a PACF je zřejmé, že rozšířením modelu o část AR(1) zůstala sezónnost zachována, proto doplníme model o sezónní část SAR(1). Budeme odhadovat model SARIMA(1,1,0)(1,0,0)$_c$.

Tab. 3.5c: Výstupní tabulka modelu SARIMA(1,1,0)(1,0,0)$_c$

```
Nonseasonal differencing of order: 1
Forecast model selected: ARIMA(1,1,0)x(1,0,0)12 with constant
Number of forecasts generated: 12
Number of periods withheld for validation: 0

            Estimation      Validation
Statistic   Period          Period
--------------------------------------------
MSE         112,619
MAE         7,91193
MAPE        0,761961
ME          -0,0864162
MPE         -0,00282616

                            ARIMA Model Summary
Parameter           Estimate        Stnd. Error     t               P-value
--------------------------------------------------------------------------------
AR(1)               -0,0453992      0,115606        -0,392706       0,695568
SAR(1)              0,737539        0,0825708       8,9322          0,000000
Mean                9,42412         3,20856         2,93718         0,004312
Constant            2,58576
--------------------------------------------------------------------------------
Backforecasting: yes
Estimated white noise variance = 114,426 with 81 degrees of freedom
Estimated white noise standard deviation = 10,697
Number of iterations: 4

Tests for Randomness of residuals
Box-Pierce Test
     Test based on first 24 autocorrelations
     Large sample test statistic = 22,1613
     P-value = 0,450295
```

Rezidua a reziduální periodogram modelu SARIMA(1,1,0)(1,0,0)$_c$ jsou obdobná jako u modelu ARIMA(1,1,0)$_c$.

I když se hodnoty reziduální ACF a PACF (obr. 3.13l,m) snížily a hodnota "P-value" Boxova-Pierceova testu se zvýšila (0,450295), z *t*-testů parametrů modelu je zřejmé, že část AR(1) do modelu nepatří.



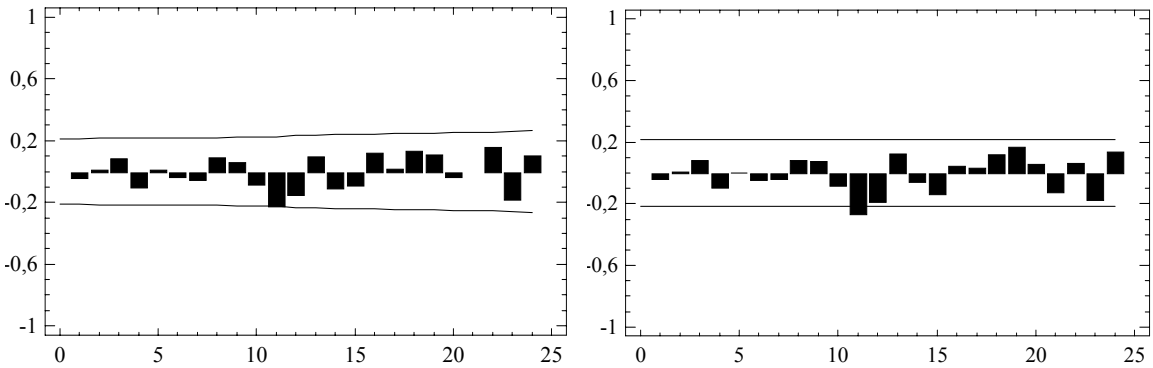
Obr. 3.13l,m: Reziduální ACF a PACF modelu SARIMA(1,1,0)(1,0,0)$_c$

Odhadneme tedy model bez AR(1), tj. model SARIMA(0,1,0)(1,0,0)$_c$.

Tab. 3.5d. Výstupní tabulka modelu SARIMA(0,1,0)(1,0,0)$_c$

```
Nonseasonal differencing of order: 1
Forecast model selected: ARIMA(0,1,0)x(1,0,0)12 with constant
Number of forecasts generated: 12
Number of periods withheld for validation: 0

            Estimation      Validation
Statistic   Period          Period
--------------------------------------------
MSE         111,572
MAE         7,89068
MAPE        0,758456
ME          -0,0466404
MPE         0,00119114

                            ARIMA Model Summary
Parameter           Estimate        Stnd. Error     t               P-value
----------------------------------------------------------------------------
SAR(1)              0,745851        0,0778021       9,58652         0,000000
Mean                9,28149         3,50721         2,64641         0,009751
Constant            2,35888
----------------------------------------------------------------------------
Backforecasting: yes
Estimated white noise variance = 112,955 with 82 degrees of freedom
Estimated white noise standard deviation = 10,628
Number of iterations: 4

Tests for Randomness of residuals
Box-Pierce Test
     Test based on first 24 autocorrelations
     Large sample test statistic = 22,4806
     P-value = 0,491417
```

Hodnoty reziduální ACF a PACF (obr. 3.13n,o) se již výrazně nesnížily, ale přesto lze předpokládat, že rezidua mají nesystematický charakter. Toto potvrzuje i hodnota Boxova-Pierceova testu (0,491417).



Obr. 3.13n,o: Reziduální ACF a PACF modelu SARIMA(0,1,0)(1,0,0)$_c$

Porovnáním hodnot "P-value" parametrů Φ_1 (0,000000) a μ (0,009751) s hladinou významnosti α (0,05) u provedeného t-testu z tab. 3.5d prokážeme, že všechny parametry jsou nenulové a odhadnutý model má tvar

$$(1 - 0{,}745851B^{12})(1 - B)y_t = 2{,}35888 + a_t,$$

lze jej vyjádřit také jako

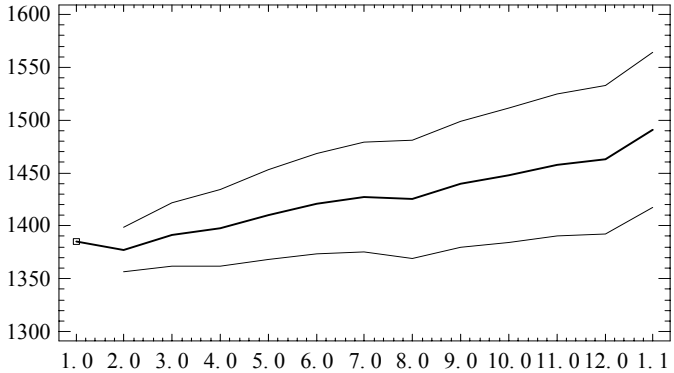
$$y_t = 2{,}35888 + y_{t-1} + 0{,}745851y_{t-12} - 0{,}745851y_{t-13} + a_t.$$

Tab. 3.5e obsahuje bodové a intervalové předpovědi měsíční časové řady na únor roku 2000 až leden roku 2001.

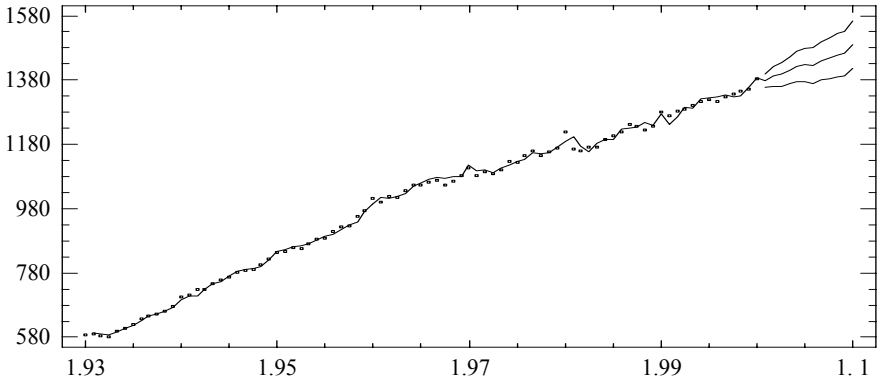
Tab. 3.5e. Předpovědi

```
Forecast Table for M2_CR
Model: ARIMA(0,1,0)x(1,0,0)12 with constant

                                    Lower 95,0%         Upper 95,0%
Period          Forecast            Limit               Limit
-----------------------------------------------------------------
2.00            1377,26             1356,12             1398,41
3.00            1391,48             1361,58             1421,38
4.00            1398,02             1361,4              1434,64
5.00            1410,6              1368,31             1452,88
6.00            1421,08             1373,81             1468,36
7.00            1427,32             1375,53             1479,11
8.00            1425,13             1369,19             1481,07
9.00            1439,35             1379,55             1499,15
10.00           1447,9              1384,47             1511,33
11.00           1457,71             1390,86             1524,57
12.00           1462,68             1392,56             1532,81
1.01            1490,85             1417,61             1564,09
```

Graf předpovědí je uveden na obr. 3.13p a graf časové řady s vyrovnanými hodnotami a předpověďmi je zachycen na obr. 3.13q.



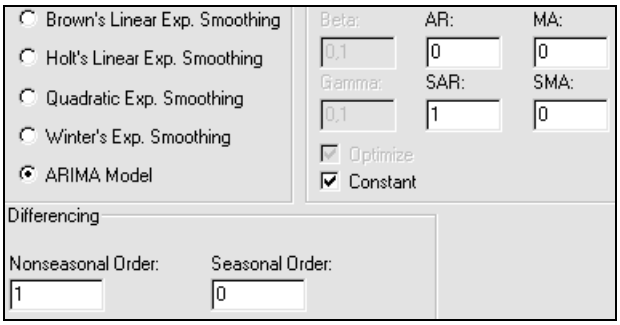
Obr. 3.13p: Předpovědi



Obr. 3.13q: Časová řada s předpověďmi

1) Ve vstupním panelu vložíme do řádku Data časovou řadu M2_CR a v části **Sampling Interval** vybereme Month(s). V políčku Starting At přepíšeme nabízenou hodnotu na 1.93 a do políčka Seasonality zadáme délku sezóny 12. Počet předpovědí 12 zadáme do políčka Number of Forecasts.

2) V doplňkovém panelu ***Model Specifications Options*** (obr. 3.7) v části **Type** zvolíme ARIMA Model a do všech políček v **Parameters and Terms** a **Differencing** zadáme hodnotu 0.
3) Z vybereme graf původní časové řady (obr. 3.13), graf autokorelační funkce a parciální autokorelační funkce (obr. 3.13a,b) a periodogram (obr. 3.13c). Z vybereme testy nahodilosti reziduí.
4) Podle grafů ACF, PACF a periodogramu původní časové řady zvolíme pro transformaci časové řady nesezónní diferenci přepsáním nuly na jedničku v políčku Nonseasonal Order v části **Differencing**. Odhadnutý model ARIMA(0,1,0)$_c$ (tab. 3.5a) se objeví automaticky ve výsledkovém okně, současně se přepočítá Boxův-Pierceův test a podle vypočítaných reziduí se překreslí všechny grafy (obr. 3.13d-g).
5) Z grafů reziduální ACF a PACF modelu ARIMA(0,1,0)$_c$ vyplývá, že bude možné rozšířit model o část AR(1). Přepíšeme proto v políčku AR nulu na jedničku. Odhadneme model ARIMA(1,1,0)$_c$ (tab. 3.5b) a zkontrolujeme překreslené grafy (obr. 3.13h-k).
6) Grafy reziduální ACF a PACF modelu ARIMA(1,1,0)$_c$ indikují, že bude vhodné přidat do modelu sezónní část SAR(1). Přepíšeme tedy v políčku SAR nulu na jedničku a odhadneme model SARIMA(1,1,0)(1,0,0)$_c$ (tab. 3.5c) a zkontrolujeme překreslené grafy (obr. 3.13l,m).
7) Z tabulky výsledků zjistíme, že odhad parametru u AR(1) je statisticky nevýznamný a proto část AR(1) z modelu vyloučíme. Do políčka AR zapíšeme nulu a odhadneme model SARIMA(0,1,0)(1,0,0)$_c$ (tab. 3.5d) a zkontrolujeme překreslené grafy (obr. 3.13n,o).



Obr. 3.13r: Výřez doplňkového panelu Model Specifications Options pro model SARIMA(0,1,0)(1,0,0)$_c$

8) Na základě zvoleného modelu SARIMA(0,1,0)(1,0,0)$_c$ vypočítáme předpovědi (tab. 3.5e). Zobrazíme graf předpovědí (obr. 3.13p) a graf původní časové řady s vyrovnanými hodnotami a předpověďmi (obr. 3.13q).

3.5.5 Cvičení

1. Pomocí Boxovy-Jenkinsovy metodologie najděte model měsíční časové řady vývozu Islandu (EXP_ISL), kterou máte k dispozici ve formě bazických indexů od ledna roku 1985 do prosince roku 1994 (průměr roku 1985 = 1). Na základě tohoto modelu vypočítejte předpovědi na 2 roky.
2. Sezónní časovou řadu vývozu Islandu (EXP_ISL) sezónně očistěte. Pro takto upravenou časovou řadu zvolte vhodný model Boxovy-Jenkinsovy metodologie a vypočítejte předpovědi na 1 rok.
3. Pro měsíční sezónně očištěnou časovou řadu indexu průmyslové produkce USA (IPP_USA), kterou máte k dispozici ve formě bazických indexů od ledna roku 1978 do března roku 1988 (1972 = 100) najděte vhodný ARIMA model a na jeho základě vypočítejte předpovědi na 1 rok.
4. Máte k dispozici čtvrtletní časovou řadu množství oběživa Švýcarska (O_SV) ve formě bazických indexů od I. čtvtletí roku 1980 do I. čtvrtletí roku 1999 (průměr roku 1980 = 1). Zvolte pro tuto časovou řadu vhodný model Boxovy-Jenkinsovy metodologie a vypočítejte předpovědi na 2 roky.
5. Rozhodněte, který z modelů SARIMA(2,1,0)(1,0,0), SARIMA(1,1,0)(1,0,0) a SARIMA(1,1,0)(1,0,1) nejlépe popisuje časovou řadu vývozu Nového Zélandu (EXP_NZ), kterou máte k dispozici ve formě bazických indexů od ledna roku 1985 do prosince roku 1994 (průměr roku 1985 = 1).

PŘÍLOHY

PŘÍLOHA A – POUŽITÉ ČASOVÉ ŘADY

Roční časové řady:

Označení	Popis	Kalendář
HDP_ARG	Hrubý domácí produkt Argentiny, bazický index 1995 = 100	1951 - 1998
HDP_CR	Hrubý domácí produkt České republiky, mld. Kč (srovnatelné ceny)	1994 – 2000
HDP_VB	Hrubý domácí produkt Velké Británie, bazický index 1995 = 100	1960 – 1997
Obyv_SR	Střední stav obyvatel Slovenska	1924 – 1997
PracStav_SR	Podíl pracujících ve stavebnictví na Slovensku v %	1987 – 1997
ZivNar_CR	Počet živě narozených dětí v České republice	1920 – 1999

Čtvrtletní časové řady:

HDP_c_CR	Hrubý domácí produkt České republiky, mld. Kč (srovnatelné ceny)	I/1994 – IV/2000
HDP_RAK	Hrubý domácí produkt Rakouska, v mld. ATS	I/1964 - IV/1998
O_SV	Množství peněz v oběhu Švýcarska, bazický index, průměr r. 1980 = 1	I/1980 – I/1999
SA_PodNez_SR	Podíl nezaměstaných 30 - 34 letých na celkové nezaměstanosti Slovenska v % (sezónně očištěno)	I/1994 – IV/2000
SA_SpEn_SR	Spotřeba elektrické energie ve firmě Stella v kWh (sezónně očištěno)	I/1995 – IV/1999

Měsíční časové řady:

EXP_ISL	Vývoz Islandu, bazický index, průměr r. 1985 = 1	1/1985 – 12/1994
EXP_NZ	Vývoz Nového Zélandu, bazický index, průměr r. 1985 = 1	1/1985 – 12/1994
IPP_USA	Index průmyslové produkce USA (sezónně očištěno)	1/1978 - 3/1988
M2_CR	Peněžní aktiva M2 České republiky mld. Kč	1/1993 - 1/2000
NezGym_SR	Počet nezaměstnaných absolventů gymnázií na Slovensku	1/1993 – 12/1996
NezNov_CR	Počet nově registrovaných uchazečů o zaměstnání na úřadech práce v České republice	1/1994 – 12/2000
NezNov_SR	Počet nově registrovaných uchazečů o zaměstnání na úřadech práce na Slovensku	1/1994 – 7/2000
NezOdr_SR	Počet odregistrovaných uchazečů o zaměstnání na úřadech práce na Slovensku	1/1993 – 12/1999
SA_NezOdr_SR	Počet odregistrovaných uchazečů o zaměstnání na úřadech práce na Slovensku (sezónně očištěno)	1/1993 – 12/1999

Použité časové řady jsou v elektronické podobě v souboru Data.sf3 a Data.xls k dispozici na internetu na adresách http://nb.vse.cz/~arltova/hlavni/vyuka/data/crsbirka02/data.sf3 (resp. .xls) a http://nb.vse.cz/~arlt/hlavni/vyuka/data/crsbirka02/data.sf3 (resp. .xls).

Roční časové řady

	HDP_ARG	HDP_CR	HDP_VB	PracStav_SR
1951	39			
1952	34			
1953	39			
1954	39			
1955	44			
1956	44			
1957	44			
1958	49			
1959	44			
1960	49		45	
1961	54		46	
1962	54		47	
1963	49		49	
1964	48		51	
1965	54		53	
1966	59		54	
1967	59		55	
1968	61		57	
1969	64		58	
1970	66		59	
1971	68		61	
1972	70		63	
1973	72		67	
1974	76		66	
1975	76		66	
1976	76		68	
1977	80		69	
1978	78		72	
1979	83		74	
1980	85		72	
1981	80		71	
1982	77		72	
1983	81		75	
1984	82		77	
1985	76		80	
1986	82		83	
1987	84		87	10,4
1988	82		91	10,2
1989	77		93	10,3
1990	75		94	10,2
1991	83		92	11,2
1992	91		91	9,1
1993	96		93	8,2
1994	104	1303,6	97	7,6
1995	100	1381,1	100	7,2
1996	105	1447,7	102	7,5
1997	114	1432,8	106	7,5
1998	119	1401,3		
1999		1390,6		
2000		1433,8		

	Obyv_SR	ZivNar_CR			Obyv_SR	ZivNar_CR
1920		244668		1960	3994270	128879
1921		257281		1961	4191977	131019
1922		248728		1962	4238056	133557
1923		241230		1963	4282865	148840
1924	3135200	228894		1964	4327949	154420
1925	3167600	225555		1965	4373595	147438
1926	3201500	219802		1966	4413853	141162
1927	3231700	208711		1967	4450880	138448
1928	3258128	208942		1968	4483656	137437
1929	3285754	203064		1969	4518773	143165
1930	3315459	207224		1970	4528459	147865
1931	3350158	196214		1971	4559341	154180
1932	3388339	190397		1972	4596330	163661
1933	3424402	176201		1973	4640673	181750
1934	3455922	171042		1974	4691014	188015
1935	3485959	170052		1975	4739301	190776
1936	3514546	169124		1976	4789452	187378
1937	3540175	170251		1977	4840819	181763
1938	3725558	185623		1978	4891673	178901
1939	3577010	192344		1979	4940223	172112
1940	3553461	209432		1980	4984331	153801
1941	3541627	208913		1981	5017032	144438
1942	3522982	215259		1982	5054770	141738
1943	3503154	225379		1983	5091537	137431
1944	3483659	230183		1984	5127719	136941
1945	3459058	225025		1985	5161789	135881
1946	3392493	210454		1986	5192789	133356
1947	3398671	206745		1987	5223609	130921
1948	3445881	197837		1988	5251120	132667
1949	3446781	185484		1989	5276186	128356
1950	3463446	188341		1990	5297774	130564
1951	3508698	185570		1991	5283404	129354
1952	3558137	180143		1992	5306539	121705
1953	3598761	172547		1993	5324632	121025
1954	3661437	168402		1994	5347413	106579
1955	3726601	165874		1995	5363676	96097
1956	3787111	162509		1996	5373810	90446
1957	3844277	155429		1997	5383214	90657
1958	3899751	141762		1998		90535
1959	3946039	128982		1999		89471

Čtvrtletní časové řady

	HDP_c_CR						
	1994	1995	1996	1997	1998	1999	2000
I.	302,2	319,9	339,1	340,4	336,8	324,2	338,2
II.	321,8	343,0	360,7	357,6	351	348,1	355,5
III.	345,2	367,9	386,2	378,0	368,6	370,0	378,2
IV.	334,4	350,3	361,7	356,8	344,9	348,3	361,9

	HDP_RAK								
	1964	1965	1966	1967	1968	1969	1970	1971	1972
I	49,00	52,09	58,47	62,05	66,65	72,15	81,48	93,16	103,65
II	54,48	59,35	64,45	69,89	74,27	81,28	92,00	100,69	113,55
III	60,38	66,58	71,73	74,89	81,30	89,25	98,88	112,72	126,59
IV	62,88	68,48	73,88	78,78	84,62	92,32	103,52	113,05	135,76
	1973	1974	1975	1976	1977	1978	1979	1980	1981
I	118,03	138,11	145,30	162,215	182,958	193,72	213,933	232,822	244,34
II	130,35	148,62	155,67	178,198	199,244	211,772	229,902	246,418	264,012
III	142,21	163,78	173,26	197,244	214,793	227,578	246,668	265,544	282,36
IV	152,86	168,06	181,89	204,442	223,949	233,752	255,388	271,358	291,026
	1982	1983	1984	1985	1986	1987	1988	1989	1990
I	263,593	278,854	298,015	309,382	323,759	336,697	353,899	381,938	415,196
II	284,933	302,445	316,203	333,596	354,937	368,61	385,556	414,507	448,077
III	302,383	321,905	335,709	357,938	373,47	386,474	403,921	426,297	462,656
IV	310,256	334,176	349,081	368,179	386,878	402,344	422,395	453,935	487,553
	1991	1992	1993	1994	1995	1996	1997	1998	
I	445,47	472,80	486,68	513,144	536,706	555,353	580,564	609,73	
II	479,80	513,005	529,867	555,685	578,453	603,855	620,933	654,00	
III	498,508	525,665	543,155	573,066	594,374	612,92	639,302	666,70	
IV	522,044	545,802	565,641	596,043	619,206	642,512	673,568	692,20	

	O_SV									
	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989
I	0,983	0,992	0,983	1,045	1,041	1,075	1,101	1,108	1,170	1,202
II	0,981	0,997	1,006	1,038	1,061	1,077	1,084	1,121	1,170	1,203
III	0,970	0,974	1,006	1,027	1,051	1,052	1,072	1,113	1,170	1,187
IV	1,066	1,037	1,089	1,103	1,169	1,145	1,196	1,212	1,284	1,295
	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999
I	1,178	1,222	1,212	1,227	1,271	1,259	1,293	1,347	1,332	1,408
II	1,181	1,226	1,221	1,230	1,261	1,266	1,301	1,324	1,349	
III	1,164	1,206	1,194	1,217	1,243	1,251	1,275	1,297	1,336	
IV	1,318	1,311	1,315	1,315	1,368	1,384	1,451	1,439	1,485	

	SA_PodNez_SR						
	1994	1995	1996	1997	1998	1999	2000
I.	14,273	14,273	13,8846	13,2049	12,4282	11,4572	10,9717
II.	14,8985	14,4958	13,8918	13,1872	12,1805	10,8719	10,7712
III.	14,5173	14,1054	13,4877	12,8699	12,0462	11,1196	10,6048
IV.	14,3222	13,9243	13,3276	12,7308	11,8357	11,04	10,6422

	SA_SpEn_SR				
	1995	1996	1997	1998	1999
I.	782,99	698,49	645,63	810,8	907,07
II.	622,34	603,51	764,05	845,31	945,4
III.	677,23	673,55	726,99	869,81	899,29
IV.	649	782,92	821,04	728,32	994,1

Měsíční časové řady

	EXP_ISL									
	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994
1	0,825	0,853	0,725	0,830	1,290	1,139	1,461	1,248	1,264	1,457
2	0,753	1,147	1,169	1,299	1,575	1,617	1,660	1,704	1,461	1,532
3	0,849	1,045	1,754	2,166	1,881	2,068	2,545	2,199	2,010	2,466
4	0,938	1,641	2,185	2,000	1,908	1,927	1,754	1,662	1,797	2,026
5	0,997	1,483	1,834	1,875	1,834	2,345	2,217	2,449	1,682	1,760
6	1,135	1,355	2,073	1,997	1,913	2,198	2,065	1,680	1,784	2,048
7	1,197	1,448	2,052	1,692	1,687	2,088	1,923	1,967	1,729	1,762
8	1,159	1,472	1,484	1,522	1,484	1,910	1,891	2,084	1,485	1,808
9	0,882	1,409	1,661	1,971	1,400	2,091	1,715	2,051	1,915	2,265
10	0,956	1,483	1,679	1,584	1,593	1,889	1,883	1,836	2,011	2,042
11	0,920	1,094	1,651	1,895	1,737	1,882	1,950	1,838	1,570	2,399
12	1,389	1,749	1,978	2,132	2,036	2,281	1,754	1,788	1,895	2,334

	EXP_NZ									
	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994
1	0,946	0,827	0,996	1,085	1,483	1,604	1,507	1,337	1,353	1,543
2	1,049	1,010	1,145	1,595	1,620	1,579	1,758	1,654	1,712	2,010
3	1,094	0,921	1,165	1,727	1,771	1,782	1,693	1,996	2,201	2,402
4	0,998	1,069	1,304	1,402	1,629	1,527	1,667	1,956	2,026	2,055
5	1,126	1,222	1,384	1,771	1,992	1,912	1,982	1,906	2,129	2,633
6	0,918	1,050	1,423	1,683	1,670	1,884	1,658	1,972	2,010	2,157
7	1,007	1,196	1,323	1,413	1,377	1,583	1,683	1,750	1,799	1,991
8	0,983	0,944	1,205	1,492	1,341	1,628	1,605	1,443	1,736	2,010
9	0,913	1,037	1,194	1,557	1,370	1,550	1,495	1,602	1,720	2,083
10	0,960	1,041	1,181	1,095	1,081	1,639	1,552	1,692	1,677	1,978
11	0,973	0,972	1,175	1,691	1,680	1,728	1,864	1,606	1,902	2,286
12	1,031	1,044	1,601	1,917	1,604	1,487	1,651	1,698	1,839	2,405

	IPP_USA										
	1978	1979	1980	1981	1982	1983	1984	1985	1986	1987	1988
1	126,5	133,0	139,7	150,4	152,5	150,0	143,4	134,7	156,2	166,2	168,8
2	125,9	132,3	138,8	152,0	152,7	151,4	140,7	137,4	158,5	165,6	169,6
3	127,6	133,3	139,2	152,5	152,6	151,8	142,9	138,1	160,0	165,7	168,4
4	128,3	135,3	140,9	153,5	152,1	152,1	141,7	140,0	160,8	166,1	
5	128,7	136,1	143,2	151,1	148,3	151,9	140,2	142,6	162,1	166,2	
6	129,7	137,0	143,9	152,7	144,0	152,7	139,2	144,4	162,8	166,2	
7	129,8	137,8	144,9	153,0	141,5	152,9	138,7	146,4	164,4	166,5	
8	130,7	138,7	146,1	153,0	140,4	153,9	138,8	149,7	165,9	166,2	
9	131,3	138,1	147,1	152,1	141,8	153,6	138,4	151,8	166,0	167,7	
10	130,6	138,5	147,8	152,7	143,9	151,6	137,3	153,8	165,0	167,6	
11	130,2	138,9	148,6	152,7	146,5	149,1	135,8	155,0	164,5	166,6	
12	131,5	139,3	149,5	152,3	148,5	146,3	134,8	155,3	165,2	167,6	

	M2_CR							
	1993	1994	1995	1996	1997	1998	1999	2000
1	588,1	704,6	845,1	1012,3	1105,8	1217,6	1280,8	1384,9
2	589,6	710,5	846,6	999,3	1083,4	1164,4	1267,4	
3	584,3	727,9	859,7	1018,9	1094,3	1160,1	1283,3	
4	581,5	729,4	856,6	1014,3	1088,4	1172,3	1288,9	
5	599,7	746	870,3	1035,9	1100,2	1172,1	1302,6	
6	606,4	759,4	886,3	1054,8	1126,5	1196,3	1313,5	
7	618,8	768,1	888,1	1053,9	1123,2	1207,8	1318,7	
8	636,6	781,3	908,3	1063,5	1145,7	1217,4	1312,6	
9	645,1	786,8	924,1	1069,5	1159,3	1240,9	1328,5	
10	651,9	790,6	926,1	1054,3	1144,5	1236,2	1336,8	
11	659,8	805,4	955,9	1065,1	1155,7	1225,9	1346,8	
12	675,3	821,9	974,0	1083,7	1169,0	1235,4	1350,3	

	NezGym_SR			
	1993	1994	1995	1996
1	3154	3756	3771	3731
2	3078	3589	3709	3590
3	3097	3474	3777	3473
4	2907	3401	3509	3399
5	2769	3268	3420	3213
6	3357	3899	3942	3585
7	4066	4473	4505	4239
8	4352	4733	4911	4581
9	4537	4500	4951	5099
10	4419	4388	4755	4906
11	4069	4079	4525	4537
12	3937	3870	4083	4417

	NezNov_CR						
	1994	1995	1996	1997	1998	1999	2000
1	55200	48000	44600	41100	50900	59600	76500
2	30000	29700	26200	26500	31200	39700	51800
3	29800	29600	25300	24200	29300	41500	56800
4	29700	28300	24300	25400	35600	45300	52200
5	26800	27200	24300	23100	30800	40200	50600
6	31400	30300	27800	27100	40100	53600	61900
7	35600	31000	29700	37900	49000	61700	65400
8	37300	31200	30100	30700	38800	54000	57500
9	47500	40300	36400	43800	66700	81500	78100
10	36400	32500	30400	34800	45400	53700	55700
11	38000	31700	30000	32800	40100	54800	56500
12	31900	27300	23200	29200	40600	53100	58200

	NezNov_SR						
	1994	1995	1996	1997	1998	1999	2000
1	33370	40330	46753	48230	49197	51041	46273
2	20767	23443	25107	24995	25887	28806	26337
3	21473	21984	21824	22158	25753	31499	26271
4	21404	21289	25580	30482	28886	32054	26356
5	20568	21554	23897	25998	32362	36672	34358
6	31615	33263	36089	34350	45906	57798	45524
7	34512	39093	41529	41223	44217	43802	41816
8	25013	27406	26218	26321	27963	27632	
9	31727	34332	38151	40252	40350	37759	
10	28446	30480	31310	34515	36038	36720	
11	29573	32398	31630	32903	37778	42138	
12	28848	31570	33965	37743	38743	36638	

	NezOdr_SR						
	1993	1994	1995	1996	1997	1998	1999
1	24370	22295	24725	27474	24374	24450	22662
2	21476	23907	25642	25568	22379	23062	21769
3	25805	27010	35294	31279	28681	30195	25981
4	30676	30832	38719	46353	42960	35671	38804
5	27380	27631	36766	38976	42648	38329	33685
6	25444	25569	33147	31647	34957	29302	27140
7	23849	24122	34997	31112	29439	30411	28885
8	24035	28898	31710	31786	25200	31463	29936
9	32527	34738	37152	40314	36222	39600	48889
10	28001	32112	41013	36824	37046	34156	37795
11	23781	26444	31609	27594	38851	22557	24606
12	18244	20372	24558	16668	18369	13428	15405

	SA_NezOdr_SR						
	1993	1994	1995	1996	1997	1998	1999
1	29914,8	27367,6	30350,5	33725,0	29919,7	30013,0	27818,1
2	27027,0	30086,3	32269,8	32176,6	28163,4	29022,9	27395,7
3	26143,5	27364,3	35757,0	31689,3	29057,2	30591,1	26321,8
4	23979,1	24101,0	30266,2	36233,6	33581,3	27883,6	30332,6
5	22974,4	23185,0	30850,1	32704,5	35785,6	32161,6	28264,8
6	25575,8	25701,5	33318,7	31810,9	35138,1	29453,8	27280,6
7	24480,9	24761,1	35924,2	31936,3	30219,0	31216,7	29650,3
8	24632,8	29616,7	32498,6	32576,5	25826,7	32245,5	30680,5
9	26151,0	27928,6	29869,4	32411,6	29121,7	31837,5	39305,7
10	23940,2	27455,0	35065,1	31483,6	31673,4	29202,6	32313,8
11	24985,5	27783,4	33210,0	28991,6	40818,8	23699,5	25852,3
12	29301,1	32718,8	39441,8	26769,9	29501,9	21566,3	24741,5

PŘÍLOHA B - KRITICKÉ HODNOTY PRO DURBINŮV-WATSONŮV TEST (α = 0,05)

n	*p* = 2		*p* = 3		*p* = 4		*p* = 5		*p* = 6	
	d	*h*	*d*	*h*	*d*	*h*	*d*	*h*	*d*	*h*
15	1,077	1,361	0,946	1,543	0,814	1,750	0,685	1,977	0,562	2,220
16	1,106	1,371	0,982	1,539	0,857	1,728	0,734	1,935	0,615	2,157
17	1,133	1,381	1,015	1,536	0,897	1,710	0,779	1,900	0,664	2,104
18	1,158	1,391	1,046	1,535	0,933	1,696	0,820	1,872	0,710	2,060
19	1,180	1,401	1,074	1,536	0,967	1,685	0,859	1,848	0,752	2,023
20	1,201	1,411	1,100	1,537	0,998	1,676	0,894	1,828	0,792	1,991
21	1,221	1,420	1,125	1,538	1,026	1,669	0,927	1,812	0,829	1,964
22	1,239	1,429	1,147	1,541	1,053	1,664	0,958	1,797	0,863	1,940
23	1,257	1,437	1,168	1,543	1,078	1,660	0,986	1,785	0,895	1,920
24	1,273	1,446	1,188	1,546	1,101	1,656	1,013	1,775	0,925	1,902
25	1,288	1,454	1,206	1,550	1,123	1,654	1,038	1,767	0,953	1,886
26	1,302	1,461	1,224	1,553	1,143	1,652	1,062	1,759	0,979	1,873
27	1,316	1,469	1,240	1,556	1,162	1,651	1,084	1,753	1,004	1,861
28	1,328	1,476	1,255	1,560	1,181	1,650	1,104	1,747	1,028	1,850
29	1,341	1,483	1,270	1,563	1,198	1,650	1,124	1,743	1,050	1,841
30	1,352	1,489	1,284	1,567	1,214	1,650	1,143	1,739	1,071	1,833
31	1,363	1,496	1,297	1,570	1,229	1,650	1,160	1,735	1,090	1,825
32	1,373	1,502	1,309	1,574	1,244	1,650	1,177	1,732	1,109	1,819
33	1,383	1,508	1,321	1,577	1,258	1,651	1,193	1,730	1,127	1,813
34	1,393	1,514	1,333	1,580	1,271	1,652	1,208	1,728	1,144	1,808
35	1,402	1,519	1,343	1,584	1,283	1,653	1,222	1,726	1,160	1,803
36	1,411	1,525	1,354	1,587	1,295	1,654	1,236	1,724	1,175	1,799
37	1,419	1,530	1,364	1,590	1,307	1,655	1,249	1,723	1,190	1,795
38	1,427	1,535	1,373	1,594	1,318	1,656	1,261	1,722	1,204	1,792
39	1,435	1,540	1,382	1,597	1,328	1,658	1,273	1,722	1,218	1,789
40	1,442	1,544	1,391	1,600	1,338	1,659	1,285	1,721	1,230	1,786
45	1,475	1,566	1,430	1,615	1,383	1,666	1,336	1,720	1,287	1,776
50	1,503	1,585	1,462	1,628	1,421	1,674	1,378	1,721	1,335	1,771
55	1,528	1,601	1,490	1,641	1,452	1,681	1,414	1,724	1,374	1,768
60	1,549	1,616	1,514	1,652	1,480	1,689	1,444	1,727	1,408	1,767
65	1,567	1,629	1,536	1,662	1,503	1,696	1,471	1,731	1,438	1,767
70	1,583	1,641	1,554	1,672	1,525	1,703	1,494	1,735	1,464	1,768
75	1,598	1,652	1,571	1,680	1,543	1,709	1,515	1,739	1,487	1,770
80	1,611	1,662	1,586	1,688	1,560	1,715	1,534	1,743	1,507	1,772
85	1,624	1,671	1,600	1,696	1,575	1,721	1,550	1,747	1,525	1,774
90	1,635	1,679	1,612	1,703	1,589	1,726	1,566	1,751	1,542	1,776
95	1,645	1,687	1,623	1,709	1,602	1,732	1,579	1,755	1,557	1,778
100	1,645	1,694	1,634	1,715	1,613	1,736	1,592	1,758	1,571	1,780

PŘÍLOHA C – PŘÍKLADY POUŽITÍ OPERÁTORŮ VE STATGRAPHICSU

Použité operátory:

AVG(proměnná)	aritmetický průměr
COUNT(začátek;konec;krok)	zadání posloupnosti hodnot
DIFF(proměnná)	diference
DROP(proměnná;*n*)	vynechání prvních *n* hodnot proměnné
DROPLAST(proměnná;*n*)	vynechání posledních *n* hodnot proměnné
EXP(*x*)	umocnění *e* (2,71828) na *x*-tou
GEOMEAN(proměnná)	geometrický průměr
LOG(*x*)	přirozený logaritmus
REP(proměnná;*n*)	opakování každé hodnoty *n*-krát
RESHAPE(proměnná;*n*)	rozšíření posloupnosti hodnot až do počtu *n* hodnot
ROWS(*n*;*m*)	vygenerování proměnné obsahující 1 v rozmezí řádků *n* až *m*, jinde 0
SQRT(*x*)	druhá odmocnina
SUM(proměnná)	součet hodnot
TAKE(proměnná;*n*)	výběr prvních *n* hodnot v proměnné

Příklady použití některých operátorů:

COUNT – zadání posloupnosti hodnot (např. vytvoření časové proměnné)

Syntaxe: **COUNT(začátek;konec;krok)**

Příklad: *Hodnoty časové proměnné* $t = 1, 2, \ldots, 10$

Zadání: COUNT(1;10;1)

Výsledek: 1 2 3 4 5 6 7 8 9 10

DROP – vynechání prvních *n* hodnot proměnné

Syntaxe: **DROP(proměnná;*n*)**

Příklad: *Hodnoty proměnné x* 1 2 3 4 5

Zadání: DROP(*x*;2)

Výsledek: 3 4 5

DROPLAST – vynechání posledních *n* hodnot proměnné

Syntaxe: **DROPLAST(proměnná;*n*)**

Příklad: *Hodnoty proměnné x* 1 2 3 4 5

Zadání: DROPLAST(*x*;2)

Výsledek: 1 2 3

REP – opakování každé hodnoty proměnné *n* krát

Syntaxe: **REP(proměnná;*n*)**

Příklad:	*Hodnoty proměnné x*	1 2 3 4 5
	Zadání:	REP(*x*;2)
	Výsledek:	1 1 2 2 3 3 4 4 5 5

RESHAPE – rozšíření posloupnosti hodnot až do počtu *n* hodnot

Syntaxe: **RESHAPE(proměnná;*n*)**

Příklad:	*Hodnoty proměnné x*	1 2 3 4 5
	Zadání:	RESHAPE(*x*;8)
	Výsledek:	1 2 3 4 5 1 2 3

ROWS – vygenerování proměnné obsahující 1 v rozmezí řádků *n* až *m*, jinde 0 (např. vytvoření umělých nulajedničkových proměnných)

Syntaxe: **ROWS(*n*;*m*)**

Příklad:	*Hodnoty umělé proměnné*	1 0 0 0 0
	Zadání:	TAKE(ROWS(1;1);5)
	Výsledek:	1 0 0 0 0
Příklad:	*Hodnoty umělé proměnné*	0 1 0 0 0
	Zadání:	TAKE(ROWS(2;2);5)
	Výsledek:	0 1 0 0 0

LITERATURA

ARLT, J. (1999): Moderní metody modelování ekonomických časových řad. Grada Publishing, Praha.

ARLT, J. - ARLTOVÁ, M. (1997): Příklady z analýzy ekonomických časových řad, skripta VŠE Praha.

ARLTOVÁ, M. - KOZÁK, J. - MATUŠŮ, M. (1999): STATGRAPHICS for Windows (Zadávání úloh), skripta VŠE Praha.

BAKYTOVÁ, H. (1990): Teória štatistiky. ES VŠE v Bratislave.

BOWERMAN, B. L. - O´CONNELL, R. T. (1987): Time Series Forecasting. Boston, Duxbury Press.

BROWN, R. G. (1959). Statistical forecasting for inventory control, New York.

BROWN, R. G. – MEYER, R. F. (1961): The fundamental theory of exponential smoothing. Operations Research, 9, str. 673-685.

CIPRA, T. (1986): Analýza časových řad s aplikacemi v ekonomii, SNTL Praha.

FREES, E. (1996): Data Analysis Using Regression Models. The Business Perspective. Prentice Hall, Englewood Cliffs, New Jersey.

GRANGER, C. W. J. - NEWBOLD, P. (1977): Forecasting Economic Time series. Academic Press, New York.

CHAJDIAK, J. – KOMORNÍK, J. – KOMORNÍKOVÁ, M. (1999). Štatistické metódy, STATIS, Bratislava.

CHAJDIAK, J. – RUBLÍKOVÁ, E. – GUDÁBA, M. (1994): Štatistické metódy v praxi. STATIS, Bratislava.

CHAJDIAK, J. a kol. (1998): Príručka pre užívateľov systému STATGRAPHICS. 2. Vydanie, STATIS, Bratislava.

International Financial Statistics 6/1999, International Monetary Fund, Washington, D. C., USA.

KENDALL, M. (1976): Time-series. London, Griffin.

KOZÁK, J. - HINDLS, R. - ARLT, J. (1994): Úvod do analýzy ekonomických časových řad, skripta VŠE Praha.

KUBANOVÁ, J. - LINDA, B. (1999): Porovnání vybraných metod odhadu parametrú Gompertzovy křivky. Sborník VI. medzinárodnej konferencie, Kvantitatívne metódy v ekonómii a podnikaní, FHI EU Bratislava.

ЛУКАШИН, Ю. П. (1979): Адаптивные методы краткосрочного прогнозирования, Statistika, Moskva.

MONTGOMERY, D. C. - JOHNSON, L. A. - GARDINER, J. S. (1990): Forecasting Time Series Analysis. Mc Graw-Hill Inc.

PINDYCK, R. S. – RUBINFELD, D. L. (1987): Econometric Models and Economic Forecasts. McGraw- Hill Inc.

RUBLÍKOVÁ, E. (1989). Prognostická štatistika - Adaptívne metódy, ES VŠE Bratislava.

RUBLÍKOVÁ, E. - ARLT, J. - ARLTOVÁ, M. - LIBIČOVÁ, L. (2001): Analýza časových radov - Zbierka príkladov, skripta Ekonomické university v Bratislavě, Ekonóm.

RUBLÍKOVÁ, E. – PACÁKOVÁ, V. (2000): Štatistické modely v analýzach trhu práce, Ekonóm, Bratislava.

SPIEGEL, M. R. (1991): Theory and Problems of Statistics. McGraw-Hill Inc.

Statistická ročenka ČR

Statistická ročenka SR

Statistické přehledy. Statistický údajový měsíčník České republiky 2/96, ČSÚ.

Statistika, časopis Českého statistického úřadu 12/95, 7/96.

WEI, W. W. S. (1990): Time Series Analysis, Univariate and Multivariate Methods, Addison–Wesley Publ. Comp., Inc.