MASARYKOVA UNIVERZITA • EKONOMICKO-SPRÁVNÍ FAKULTA

EKONOMICKO-MATEMATICKÉ METODY II B
podklady pro cvičení

HANA PYTELOVÁ

BRNO 2005

Obsah

1. Základy práce v Matlabu

1.1 Základní příkazy

- Do pracovního adresáře se přepneme tak, že na horní liště vybereme symbol ... a následně zvolíme adresář, kde budeme pracovat. Nebo využijeme okno vlevo nahoře, záložka `Current Directory`, kde si adresář vybereme.
- Obsah aktuálního adresáře vypíšeme pomocí příkazu `dir`
- Pokud chceme, aby se veškeré v Matlabu námi napsané příkazy včetně výpisů na obrazovku uložily do souboru `nazev.txt`, na začátku práce napíšeme `diary nazev.txt`, na konci práce napíšeme `diary off`.
- Vše co je napsáno na řádku za symbolem % se ignoruje, používá se k psaní komentářů.
- Ukončení programu provedeme příkazem `quit` nebo `exit` nebo okno zavřeme křížkem vpravo nahoře.
- Pokud něco nevíme (např. k čemu je funkce sin), použijeme nápovědu, tj. napíšeme `help sin`.
- Příkaz `clear all` smaže všechny proměnné. Příkaz `save` slouží k uložení proměnných. Příkaz `load` slouží k načtení proměnných uložených dříve pomocí příkazu `save`.
- Příkaz `close all` zavře všechny obrázky.
- Příkaz `clc` smaže příkazové okno.
- Příkaz `format bank` bude zobrazovat čísla na 2 desetinná místa, `format short` na 4 desetinná místa, atd. - viz `help format`.

1.2 Matice, vektory

1.2.1 Vytvoření

Prvky matice píšeme do hranatých závorek, prvky na řádku oddělujeme mezerou nebo čárkou, jednotlivé řádky oddělíme středníkem nebo klávesou enter. Tj. níže uvedenou matici

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}$$

vytvoříme např.

`A=[1,2,3;4,5,6;7,8,9]` nebo
`A=[1 2 3;4 5 6;7 8 9]` nebo

```
A=[1 2 3
   4 5 6
   7 8 9]
```

Podobně lze vytvářet vektory. Speciální vektory lze také tvořit výčtem pomocí symbolu `:` uvedeného mezi první a poslední složkou (viz vektor `v1`). Automaticky nastavený krok (tj. vzdálenost složek) je jedna. Chceme-li krok jiný, uvedeme jej mezi dvojtečky mezi první a poslední složkou (viz vektor `v2`). Konstanta je matice řádu `1x1`. Níže uvedeným způsobem vytvoříme řádkové vektory, pokud chceme sloupce, použijeme symbolu apostrof, tj. `'`).

$$v^T = \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix} \qquad v_1^T = \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{pmatrix} \qquad v_2^T = \begin{pmatrix} 2 \\ 1.8 \\ 1.6 \\ 1.4 \\ 1.2 \\ 1 \end{pmatrix} \qquad a = 1$$

`v = [1 2 3 4]` resp. sloupec `v = [1 2 3 4]'`
`v1= (1:6)` resp. sloupec `v1= (1:6)'`
`v2= (2:-0.2:1)` resp. sloupec `v2= (2:-0.2:1)'`
`a=1`

K potlačení výpisu na obrazovku slouží symbol `;` uvedený za příkazem. Chceme-li vidět, co je obsahem nějaké proměnné, napíšeme její název (do příkazového okna) + enter.

K výpisu všech proměnných slouží příkaz `who`, podrobnější výpis (včetně velikostí) příkaz `whos`.

Vlevo nahoře je okno se záložkou `Workspace` s informacemi o proměnných. Vlevo dole je okno `command history` s historií použitých příkazů. Dvojtým kliknutím lze příkaz provést.

1.2.2 Speciální typy matic

V Matlabu jsou k dispozici příkazy pro tvorbu speciálních typů matic. Příkaz `eye` vytvoří jednotkovou matici, příkaz `ones` vytvoří matici ze samých jedniček, příkaz `zeros` vytváří nulovou matici. Zadáme-li za příkaz dva parametry, vytvoří se matice obdélníková, zadáme-li jeden, bude čtvercová daného řádu. Tedy např. matice

$$E = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \qquad Z = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

$$J1 = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \quad J2 = \begin{pmatrix} 1 & 1 \end{pmatrix} \quad J3 = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$$

lze vytvořit pomocí příkazů

```
E=eye(3)
Z=zeros(2,3)
J1=ones(2)
J2=ones(1,2)
J3=ones(size(E))
```

(vytvoří jedničkovou matici stejně velkou jako matice E).

1.2.3 Operace s prvky matic a operace s maticemi

Z matice A chceme vybrat prvek na 2. řádku v 1. sloupci a uložit do proměnné p.

```
p=A(2,1)
```

Z matice A chceme vyberat řádky 2 až 3 a sloupce 1 a 3 a uložit do matice S.

```
S=A(2:3,[1 3])
```

Do matice A na pozici $(1,1)$ uložíme prvek, který je v matici A na pozici $(2,2)$.

```
A(1,1)=A(2,2)
```

Do matice A na pozici $(1,1)$ uložíme 1.

```
A(1,1)=1
```

Vybereme 1. sloupec matice A.

```
A(:,1)
```

Vybereme 1. řádek matice A.

```
A(1,:)
```

Do proměnné x uložíme 1. řádek matice A.

```
x=A(1,:)
```

Vektory resp. matice stejných rozměrů lze sčítat a odčítat. Např. vektory x a z lze sečíst (oba jsou to řádky), ale vektory x a y sečíst nelze (y je sloupec). Vektory resp. matice lze násobit skalárem, tj. číslem. Dvě matice resp. vektory lze násobit pomocí operátoru hvězdička, tj. $*$ právě když první matice má tolik sloupců, kolik má druhá matice řádků. Pro výpočet inverzní matice slouží příkaz `inv`.

```
y=x'
z=ones(1,3)
w=(x+z)
B=A'
C=A*B
inv(C)
```

Chceme-li např. násobit navzájem si odpovídající prvky matic, nebo umocnit jednotlivé prvky matice, před operátor násobení $*$ resp. umocnění $\hat{}$ (stříška) napíšeme tečku, tj. např. matice $D = M * M$, kdežto matice $D2$ vznikla z M umocněním každého jednotlivého prvku na druhou. Vektory x a z lze násobit po složkách (jinak ne).

$$M = \begin{pmatrix} 0 & 1 \\ 2 & 3 \end{pmatrix} \quad D = \begin{pmatrix} 2 & 3 \\ 6 & 11 \end{pmatrix} \quad D2 = \begin{pmatrix} 0 & 0 \\ 4 & 9 \end{pmatrix}$$

```
M=[0 1;2 3]
D=M^2
D2=M.^2
x.*z
```

Ke zjištění délky vektoru použijeme příkaz `length`, rozměry matice vypíše příkaz `size`.

`length(x)`
`size(A)`
`[m,n]=size(A)` (rozměry matice se uloží do proměnných m a n).

1.3 Funkce

Matlab umí řadu funkcí. Za název funkce píšeme do kulatých závorek argument funkce. Podrobnosti viz nápověda. Pozor na desetinnou čárku, v Matlabu se píše tečka!!!

`sin(x)` – sinus
`cos(x)` – kosinus
`abs(x)` – absolutní hodnota
`exp(x)` – provede operaci e^x
`log(x)` – přirozený logaritmus
`round(x)` – zaokrouhlení
`sqrt(x)` – odmocnina
`sum(v)` – součet prvků vektoru
`cumsum(v)` – postupné kumulativní součty prvků vektoru
`prod(v)` – součin prvků vektoru
`cumprod(v)` – postupné kumulativní součiny prvků vektoru
`mean(v)` – průměr složek vektoru
`max(v)` – maximum
`min(v)` – minimum
`sort(v)` – setřídí prvky vektoru podle velikosti
`length(v)` – délka vektoru
`size(A)` – rozměry matice
`diag(A)` – diagonála matice
`inv(A)` – inverzní matice
`pi` – π

Funkce, operace, názvy proměnných lze též použít při tvorbě matice, např.

```
B=[a 2+3 1/2 sin(pi/2) sqrt(3) 5.7]
```

1.4 Cyklus FOR

Pro počítací proměnnou od ... do ... proveď příkazy ...

Příklad : s pomocí cyklu for vytvoříme vektor délky 6 ze samých trojek. Za příkazem nezapomenout středník pro potlačení výpisu v průběhu cyklu, jinak se to vypíše šest krát na obrazovku!!! Krok počítací proměnné může být libovolné číslo, tj. i záporné, pozor, pokud přiřazujeme něco na k-tou pozici, musí být k celé číslo!!!

```
for k=1:6
 y(k)=3;
end
```

$$y = \begin{pmatrix} 3 & 3 & 3 & 3 & 3 & 3 \end{pmatrix}$$

Příklad : pomocí cyklu for vytvoříme vektor délky 9, jehož každá lichá souřadnice bude mocninou své pozice, sudé souřadnice budou nulové. Vyjde nám řádek w.

```
for k=1:2:9
 w(k)=k^2;
end
```

$$w = \begin{pmatrix} 1 & 0 & 9 & 0 & 25 & 0 & 49 & 0 & 81 & 0 \end{pmatrix}$$

Pokud bychom chtěli vyrobit sloupec s:

```
for k=1:2:10
 s(k,:)=k^2;
end
```

Příklad : pomocí cyklu for vytvoříme vektor h, který vznikne součtem vektorů f a g uvedených níže tak, že bereme jednotlivé složky vektoru f odpředu a složky g odzadu.

```
f=(1:0.1:1.5);
g=(20:-2:10);
n=length(g);
for i=1:n
    h(i)=f(i)+g(n-i+1);
end;
```

$$f = \begin{pmatrix} 1 & 1.1 & 1.2 & 1.3 & 1.4 & 1.5 \end{pmatrix} \quad g = \begin{pmatrix} 20 & 18 & 16 & 14 & 12 & 10 \end{pmatrix}$$

$$h = \begin{pmatrix} 11 & 13.1 & 15.2 & 17.3 & 19.4 & 21.5 \end{pmatrix}$$

1.5 Podmínky

Jestliže (`if`) platí podmínka, tak proveď příkazy, (pokud neplatí, nic nedělej), konec (`end`).

```
m=1
n=3
if m<n
  m=m+1
end
```

Podmínka s větvením: `if` platí podmínka, tak proveď příkazy 1, , `else` proveď udělej příkazy 2. Příkaz `disp` vypíše na obrazovku text uvedený v apostrofech.

```
if m~=n
  m=n
  disp('nastala prvni moznost')
else
  m=10
  disp('nastala druha moznost')
end
```

1.6 Krerslení obrázků

Na kreslení obrázků slouží příkaz `plot`. Prázdný obrázek vyvoláme příkazem `figure`. Příkaz `plot(x,y)` vykreslí na vodorovnou osu hodnoty vektoru x a na svislou osu nanese hodnoty vektoru y. Popisky obrázku vytvoříme pomocí příkazů `title`, `xlabel`, `ylabel`, `legend`. Text píšeme mezi apostrofy do kulatých závorek.

```
x=0:0.1:5*pi;
y=2*x;
figure; plot(x,y);
```

Vykreslení další čáry smaže čáru předchozí. Pokud chceme vykreslit více čar do jednoho obrázku, vykreslíme je naráz jedním příkazem nebo užijeme příkazu `hold on`.

```
s=sin(x);
c=cos(x);
figure;
plot(x,s)
plot(x,c);
```
(překreslí se první čára druhou)

```
figure;
plot(x,s,x,c)
```
(obě čáry v 1 obrázku, každá čára bude jinou barvou – viz obrázek 1.1)

```
figure;
plot (x,s);
hold on;
plot(x,c)
```
(obojí v 1 obrázku, ale obě čáry budou stejné)

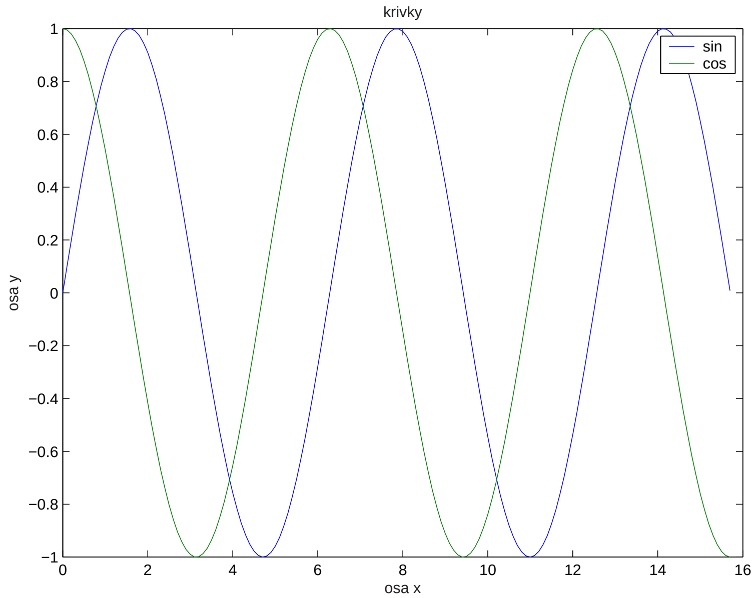
```
figure;
plot(x,[s;c])
```
([s;c] jsou dva řádky), nebo
```
plot(x,[s',c'])
```
([s' c'] jsou dva sloupce).

Nastavení barvy čáry a typu čáry – viz `help plot`. Píše se mezi apostrofy odděluje se od vykreslovaných dat čárkou – viz níže. Barvy: r, g, b, k, y, c, m (red, green, blue, black, yellow, cyan, magenta). Typy čar např.: . o x + * : -. – .- atd.

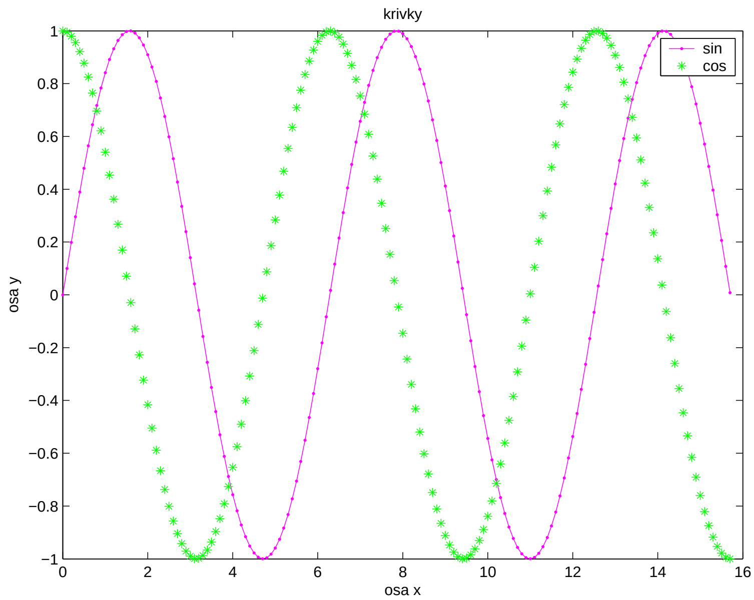
```
figure;
plot(x,s,'m.-',x,c,'g*')
title('krivky')
xlabel('osa x')
ylabel('osa y')
legend('sin','cos')
```
– viz obrázek 1.2.

1.7 Tvorba m-file

Příkazy lze psát také za sebe do matlabovského souboru (m-file) a pak je spustit naráz. Vytvoříme jej `File\New\m-file`, uložíme jako `nazev.m`. Spustíme pomocí `F5` nebo `Debug\Run` nebo pomocí tlačítka s šipkou na horní liště.



Obrázek 1.1: Obrázek z Matlabu – automatické nastavení typů čar



Obrázek 1.2: Obrázek z Matlabu – ruční nastavení typů čar

2. Odvozené ukazatele

Mějme dánu řadu přirozených čísel $\{y_t\}_{t=1}^{n}$. Pak můžeme definovat následující odvozené ukazatele:

- přírůstky se stálým základem $Dy_t = y_t - y_1$ (1)
- přírůstky s proměnlivým základem (*dynamika*)

  $dy_t = y_t - y_{t-1}$ (2)
- indexy se stálým základem

  $Ky_t = \frac{y_t}{y_1}$ (3)
- indexy s proměnlivým základem (koeficient růstu)

  $ky_t = \frac{y_t}{y_{t-1}}$ (4)
- *tempo růstu*

  $\frac{y_t - y_{t-1}}{y_{t-1}} = \frac{y_t}{y_{t-1}} - 1 = ky_t - 1 = \frac{dy_t}{y_{t-1}}$ (5a-d)
- *průměrná dynamika*

  $\bar{d} = \frac{1}{n-1}\sum_{t=2}^{n} dy_t =$ (6a)

  $= \frac{1}{n-1}((y_2 - y_1) + (y_3 - y_2) + \ldots + (y_n - y_{n-1})) = \frac{y_n - y_1}{n-1} = \frac{Dy_n}{n-1}$ (6b)
- průměrný koeficient růstu

  $\bar{k} = \sqrt[n-1]{\prod_{t=2}^{n} \frac{y_t}{y_{t-1}}} = \sqrt[n-1]{\frac{y_2}{y_1}\frac{y_3}{y_2}\cdots\frac{y_n}{y_{n-1}}} = \sqrt[n-1]{\frac{y_n}{y_1}}$ (7)
- *průměrné tempo růstu*

  $\bar{k} - 1 = \sqrt[n-1]{\frac{y_n}{y_1}} - 1 = \sqrt[n-1]{\prod_{t=2}^{n} \frac{y_t}{y_{t-1}}} - 1 = \sqrt[n-1]{\prod_{t=2}^{n} ky_t} - 1$ (8a-d)

Příklad : Vypočtěte dynamiku a průměrnou dynamiku produkce podniku v jednotlivých letech. Dále určete roční tempa růstu produkce, průměrné roční tempo růstu produkce a celkové tempo růstu produkce za celé uvažované období.

rok	produkce v tis. ks	dynamika	tempo růstu
1998	10		
1999	15		
2000	20		
2001	25		
2002	35		
průměr			

Používané míry inflace : Označme $P_{x,t}$ agregátní cenový index v měsíci x v roce t.

- Měsíční míra inflace $\pi^m_{x,t} = \frac{P_{x,t}}{P_{x-1,t}} - 1$
- Meziroční míra inflace $\pi^{y-y}_{x,t} = \frac{P_{x,t}}{P_{x,t-1}} - 1$
- Roční míra inflace $\pi^a_t = \frac{P_{12,t}}{P_{12,t-1}} - 1$

Máme-li k dispozici čtvrtletní údaje místo měsíčních, je postup výpočtu obdobný, vypočítat lze ovšem čtvrtletní míru inflace (podělením cenových indexů dvou po sobě následujících čtvrtletí), meziroční míru inflace (podělením cenových indexů pro dané čtvrtletí dvou po sobě následujících let), roční míru inflace (podělením cenových indexů čtvrtého čtvrtletí dvou po sobě následujících let).

Příklad s inflací
y_t = index CPI, vektor délky n
π = míra inflace, tj. tempo růstu indexu CPI (k), vektor délky $n-1$, $\pi = k - 1$

$$\text{průměrná inflace} = \sqrt[n-1]{\prod k} - 1 = \sqrt[n-1]{\prod_{i=2}^{n}(1+\pi_i)} - 1 \qquad \text{(9a-b)}$$

(inflaci v čase 1, tj. π_1, neznáme!!!)

Pozn. : dle vztahu (9) postupujeme i v případě výpočtu průměrného tempa růstu HDP. Označíme-li si tempo růstu HDP jako h, pak při výpočtu průměrného tempa růstu HDP ve vztahu (9b) nahradíme vektor π vektorem h.

Příklad : Známe inflaci za období 1-3 (π_1, π_2, π_3), pak

- průměrná inflace za dané období: $\sqrt[3]{(1+\pi_1)(1+\pi_2)(1+\pi_3)} - 1$
- celková inflace za dané období: $(1+\pi_1)(1+\pi_2)(1+\pi_3) - 1$ (10)

kde π_i není procentní hodnota, nýbrž desetinné číslo, tj. např. ne 4% ale 0.04.

Příklad Práce s odvozenými ukazateli v Matlabu je předvedena v souboru `odv_ukaz.m`. Jako cvičení spočtěte:

a) Tempo růstu nominálního HDP v roce 1996

b) Průměrné tempo růstu reálného HDP v letech 1997-1998

c) Roční inflaci CPI za rok 1998

d) Celkovou roční inflaci CPI v letech 1997-1999

e) Průměrnou roční inflaci PPI v letech 1998-1999

$$[a = 13.46\%,\ b = -0.9\%,\ c = 7.47\%,\ d = 20.65\%,\ 2.29\%]$$

3. Lineární regresní model

Lineární regresní model (LRM) vysvětluje chování veličiny y pomocí veličin x_i tak, že předpokládá lineární závislost na paramterech, tj.:

$$y_t = b_0 + b_1 x_{t1} + b_2 x_{t2} + \ldots + b_k x_{tk} + u_t \qquad t = 1, \ldots, n \tag{3.1}$$

y_t – vysvětlovaná proměnná v čase t (endogenní)
x_{ti} – i-tá vysvětlující proměnná v čase t (exogenní)
u_t – náhodná složka v čase t
b_i – i-tý parametr
n – počet pozorování
$k+1$ – počet parametrů

Maticový zápis LRM:
(totéž jako vztah (3.1) ale je to zapsáno souhrnně pro všechny časy $t = 1, \ldots, n$ naráz)

$$\boldsymbol{y} = \boldsymbol{X} \cdot \boldsymbol{b} + \boldsymbol{u} \tag{3.2}$$

$\boldsymbol{y}$ – vektor pozorování
$\boldsymbol{X}$ – matice plánu
$\boldsymbol{b}$ – vektor parametrů

kde

$$\begin{pmatrix} y_1 \\ \vdots \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & \ldots & x_{1k} \\ \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & \ldots & x_{nk} \end{pmatrix} \cdot \begin{pmatrix} b_0 \\ \vdots \\ b_k \end{pmatrix} + \begin{pmatrix} u_1 \\ \vdots \\ \vdots \\ u_n \end{pmatrix}$$

Metoda nejmenších čtverců dává takový odhad $\hat{\boldsymbol{b}}$ vektoru parametrů $\boldsymbol{b}$, aby byl minimalizován součet čtverců chyb (reziduí), tj.

$$SSE = \sum_{t=1}^{n} e_t^2 = \boldsymbol{e}'\boldsymbol{e} = (\boldsymbol{y} - \hat{\boldsymbol{y}})'(\boldsymbol{y} - \hat{\boldsymbol{y}}) \rightarrow \min \tag{3.3}$$

Odhad $\hat{\boldsymbol{b}}$ vektoru parametrů $\boldsymbol{b}$ metodou nejmenších čtverců (MNČ)
(Tento odhad lze odvodit z tzv. normálních rovnic, což jsou derivace kriteria SSE podle všech parametrů položené rovny nule.)

$$\hat{\boldsymbol{b}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y} \tag{3.4}$$

Vyrovnané hodnoty $\hat{\boldsymbol{y}}$ vektoru $\boldsymbol{y}$ (tj. ten náš odhad)

$$\hat{\boldsymbol{y}} = \boldsymbol{X}\hat{\boldsymbol{b}} \tag{3.5}$$

Rezidua neboli chyba vyrovnání, tj. rozdíl mezi skutečnými a vyrovnanými hodnotami

$$\boldsymbol{e} = \boldsymbol{y} - \hat{\boldsymbol{y}} \tag{3.6}$$

Tedy

$$\boldsymbol{y} = \boldsymbol{X}\hat{\boldsymbol{b}} + \boldsymbol{e} \tag{3.7}$$

Předpoklady pro použití MNČ

1) náhodná složka ($\boldsymbol{u}$) musí mít normální rozdělení
2) náhodná složka má v každém čase nulovou střední hodnotu, tj.
   $E(u_t) = 0,\ t = 1\ldots,n$
3) náhodná složka má konstantní rozptyl v čase = homoskedasticita
   $D(u_t) = E(u_t^2) = \sigma^2,\ t = 1\ldots,n$
4) náhodné složky různých období jsou vzájemně nekorelované (nulová kovariance), tj. $E(u_t \cdot u_{t+p}) = 0,\ p \neq 0,\ t = 1\ldots,n$ tedy
   3)+4) $\Leftrightarrow$ $var(\boldsymbol{u}) = E(\boldsymbol{u}'\boldsymbol{u}) = \sigma^2 \cdot I_n = \begin{pmatrix} \sigma^2 & \ldots & 0 \\ & \ddots & \\ 0 & \ldots & \sigma^2 \end{pmatrix}$
5) vysvětlující proměnné nejsou náhodné (jsou nestochastické), tj.
   $E(\boldsymbol{X}'\boldsymbol{u}) = 0$
6) vysvětlující proměnné jsou navzájem nezávislé a jejich počet je menší než počet pozorování, tj. $h(\boldsymbol{X}) = k + 1 \leq n$

Příklad 3.1. Konstantní trend. Metodou nejmenších čtverců odhadněte parametr $\hat{b_0}$ konstantního trendu. To znamená, že daty chceme proložit polynom stupně nula neboli konstantu, což je přímka rovnoběžná s vodorovnou osou.

$$y_t = b_0 + u_t, \quad t = 1, \ldots, n; \qquad SSE = \sum_{t=1}^{n} (y_t - \hat{b_0})^2 \rightarrow \min$$

$$\frac{\partial SSE}{\partial \hat{b_0}} = 2 \sum_{t=1}^{n} (y_t - \hat{b_0})(-1) = 0$$

$$\sum_{t=1}^{n} y_t - \sum_{t=1}^{n} \hat{b_0} = 0 \;\Rightarrow\; \sum_{t=1}^{n} y_t = n\hat{b_0} \;\Rightarrow\; \hat{b_0} = \frac{1}{n} \sum_{t=1}^{n} y_t = \overline{y}$$

Totéž ale maticově s použitím vztahu (3.4) pro odhad parametru $\hat{\boldsymbol{b}}$:

Matice plánu a vektor parametrů pro konstantní trend:

$$\boldsymbol{X} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \qquad \boldsymbol{b} = \begin{pmatrix} b_0 \end{pmatrix}$$

$$\hat{\boldsymbol{b}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y} = \left(\begin{pmatrix} 1 & \ldots & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \right)^{-1} \cdot \begin{pmatrix} 1 & \ldots & 1 \end{pmatrix} \cdot \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} =$$

$$= (n)^{-1} \cdot \sum_{t=1}^{n} y_t = \frac{1}{n} \sum_{t=1}^{n} y_t = \overline{y}$$

$[\overline{y}]$

Příklad 3.2. Lineární trend. Metodou nejmenších čtverců odhadněte parametry $\hat{b_0}$ a $\hat{b_1}$ lineárního trendu. To znamená, že daty chceme proložit polynom stupně jedna neboli přímku se směrnicí $\hat{b_1}$ posunutou na svislé ose o $\hat{b_0}$.

$$y_t = b_0 + b_1 \cdot t + u_t, \qquad SSE = \sum_{t=1}^{n} (y_t - \hat{b_0} - \hat{b_1} \cdot t)^2 \rightarrow \min$$

$$\frac{\partial SSE}{\partial \hat{b_0}} = \sum_{t=1}^{n} 2(y_t - \hat{b_0} - \hat{b_1}t)(-1) = 0$$
$$\frac{\partial SSE}{\partial \hat{b_1}} = \sum_{t=1}^{n} 2(y_t - \hat{b_0} - \hat{b_1}t)(-t) = 0$$

Z prvního vztahu vyjádříme $\hat{b_0}$ a dosadíme do vztahu druhého, z něhož pak vypočteme $\hat{b_1}$, což je vaše domácí úloha :o) !!!
Ti, kterým to přijde příliš jednoduché, odvodí totéž také maticově (jako u konstantního trendu) a obdrží samozřejmě +

Matice plánu a vektor parametrů pro lineární trend:

$$\boldsymbol{X} = \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ \vdots & \vdots \\ 1 & n \end{pmatrix} \qquad \hat{\boldsymbol{b}} = \begin{pmatrix} b_0 \\ b_1 \end{pmatrix}$$

$$\left[\hat{b_0} = \overline{y} - \hat{b_1}\bar{t} = \frac{\overline{y}\sum t^2 - \bar{t}\sum t y_t}{\sum t^2 - \bar{t}^2 n};\ \hat{b_1} = \frac{\sum t y_t - \bar{t}\sum y_t}{\sum t^2 - n\bar{t}^2}\right]$$

Příklad 3.3. Polynomiální trend. Daty chceme proložit polynom stupně k .

$$y_t = b_0 + b_1 \cdot t + b_2 \cdot t^2 + \ldots + b_k \cdot t^k + u_t \qquad t = 1, \ldots, n$$
$$SSE = \sum_{t=1}^{n} (y_t - \hat{b_0} - \hat{b_1} \cdot t^1 - \ldots - \hat{b_k} \cdot t^k)^2 \rightarrow \min$$

Matice plánu a vektor parametrů pro polynomiální trend stupně k:

$$\boldsymbol{X} = \begin{pmatrix} 1 & 1 & 1 & \ldots & 1 \\ 1 & 2 & 2^2 & \ldots & 2^k \\ 1 & 3 & 3^2 & \ldots & 3^k \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & n & n^2 & \ldots & n^k \end{pmatrix} \qquad \hat{\boldsymbol{b}} = \begin{pmatrix} b_0 \\ b_1 \\ \vdots \\ b_k \end{pmatrix}$$

4. Identifikace a verifikace modelu

koeficient determinace R^2

$$R^2 = 1 - \frac{\sum_{t=1}^{n} (y_t - \hat{y}_t)^2}{\sum_{t=1}^{n} (y_t - \overline{y})^2}$$

$R^2 \in \langle 0, 1 \rangle$, udává, kolik procent chování vysvětlované veličiny model vysvětluje, měl by vycházet blízko jedné.

statistická významnost koeficientu determinace
K testování statistické významnosti koeficientu determinace se používá tzv. F_R statistika, která má Fisherovo rozdělení s počty stupňů volnosti k a $n - k + 1$ na hladině významnosti α. Je-li její hodnota větší než kritická hodnota Fisherova rozdělení s příslušnými stupni volnosti, považujeme model za statisticky významný.

$$F_R = \frac{R^2/k}{(1 - R^2)/(n - (k + 1))} > F_{k,n-(k+1)}(\alpha)$$

korigovaný koeficient determinace $\overline{R}^2$
Koeficient determinace nikdy s přidáním nové vysvětlující proměnné neklesá, proto se někdy místo něj používá korigovaný koeficient determinace, který počet vysvětlujících proměnných zohledňuje.

$$\overline{R}^2 = R^2 - \frac{k}{n - (k + 1)}(1 - R^2) = 1 - (1 - R^2)\frac{n - 1}{n - (k + 1)}$$

statistická významnost parametrů

K testování statistické významnosti parametru slouží tzv. t-test. Je-li vypočtená hodnota $|t_i|$ větší než příslušný kvantil Studentova rozdělení, je parametr statisticky významný, neboli statisticky významně různý od nuly. Tento postup je ekvivalentni rozhodnutí na základě intervalu spolehlivosti parametru – obsahuje-li nulu, není parametr statisticky významný.

Symbol s_{b_i} značí směrodatnou odchylku parametru b_i, $t_{n-(k+1)}(\alpha)$ je příslušný kvantil Studentova rozdělení.

$$|t_i| = \frac{|b_i|}{s_{b_i}} > t_{n-(k+1)}(\alpha)$$

Interval spolehlivosti parametru b_i na hladině významnosti α je intervalem, v němž leží hodnota parametru b_i s pravděpodobností $1-\alpha$.

$$1-\alpha = P(b_i \in (\hat{b_i} - t_{n-(k+1)}(\alpha)\cdot s_{b_i};\ \hat{b_i} + t_{n-(k+1)}(\alpha)\cdot s_{b_i}))$$

kde

$$s_{b_i} = \sqrt{s^2(\boldsymbol{X}'\boldsymbol{X})^{-1}_{ii}} \qquad s^2 = \frac{\boldsymbol{e}'\boldsymbol{e}}{n-(k+1)}$$

Durbin-Watsonův test

Slouží k testování autokorelace reziduí, jejíž přítomnost může významně zkreslit směrodatné odchylky parametrů a tudíž i testy významnosti parametrů. Umožňuje však poznat pouze korelaci sousedních reziduí, tj. korelaci 1. řádu. Pro její použití je třeba alespoň 15 pozorování.

Hodnoty dw blízké nule značí pozitivní autokorelaci, naopak blízkost 4 značí korelaci negativní, kolem 2 jsou rezidua nekorelovaná. Existují také oblasti, ve kterých neumíme rozhodnout – viz tabulky.

$$dw = \frac{\sum_{t=2}^{n}(e_t - e_{t-1})^2}{\sum_{t=1}^{n} e_t^2} \in \langle 0; 4\rangle$$

Goldfeld-Quandtův test

Heteroskedasticita neboli nesplnění předpokladu konstantního rozptylu vede ke zkreslení směrodatných odchylek parametrů a tudíž i ke zkreslení testů významnosti parametrů, dále způsobuje problémy při konstrukci intervalů spolehlivosti.

K testování heteroskedasticity slouží např. Goldfeld-Quandtův test. Setřídíme rezidua tak, jak předpokládáme, že roste rozptyl náhodné složky (u polynomiálních modelů předpokládáme, že roste v čase), několik ($M < \frac{n}{4}$) prostředních hodnot vypustíme a vypočteme součet čtverců obou úseků, větší hodnotu podělíme menší a porovnáme s příslušným kvantilem Fisherova rozdělení s počtem stupňů volnosti ν, ν, kde $\nu = \frac{n-M}{2} - (k+1)$.

$$F_{2,1} = \frac{SSE_2}{SSE_1} = \frac{\sum e_{2,t}^2}{\sum e_{1,t}^2} > F_{\nu,\nu}(\alpha) \;\Rightarrow\; \text{heteroskedasticita}$$

Intervaly spolehlivosti vyrovnání

$$1 - \alpha = P(y_t \in (\hat{y}_t - t_{n-(k+1)}(\alpha) \cdot s \cdot f_t;\; \hat{y}_t + t_{n-(k+1)}(\alpha) \cdot s \cdot f_t))$$

kde

$$f_t = \sqrt{1 + \boldsymbol{x}_t \cdot (\boldsymbol{X}'\boldsymbol{X})^{-1} \cdot \boldsymbol{x}_t'} \qquad s^2 = \frac{\sum e_t^2}{n-(k+1)} = \frac{\boldsymbol{e}'\boldsymbol{e}}{n-(k+1)}$$

$$\boldsymbol{x}_t = \boldsymbol{X}(t,:) \quad \text{tj. t-tý řádek matice plánu}$$

Předpověď a intervaly spolehlivosti předpovědi

$$1 - \alpha = P(yp_t \in (\hat{yp}_t - t_{n-(k+1)}(\alpha) \cdot s \cdot f_t;\; \hat{yp}_t + t_{n-(k+1)}(\alpha) \cdot s \cdot f_t))$$

kde

$$f_t = \sqrt{1 + \boldsymbol{xp}_t \cdot (\boldsymbol{X}'\boldsymbol{X})^{-1} \cdot \boldsymbol{xp}_t'} \qquad s^2 = \frac{\sum e_t^2}{n-(k+1)} = \frac{\boldsymbol{e}'\boldsymbol{e}}{n-(k+1)}$$

$$\boldsymbol{xp}_t = \boldsymbol{XP}(t,:) \quad \text{tj. t-tý řádek předpovědní matice plánu}$$

Příklad Na základě dat uvedených v souboru `lintrend.m` se pokuste analyzovat závislost spotřeby elektrické energie na čase nejprve pomocí lineárního trendu. Jako domácí cvičení ručně spočtěte koeficient determinace, korigovaný koeficient determinace, hodnotu FR statistiky pro testování statistické významnosti koeficientu determinace a odpovídající kritickou hodnotu Fisherova rozdělení, dále odhad rozptylu reziduí s^2 a odhady směrodatných odchylek parametrů modelu s_{b_i}. Výsledky zkontrolujte s výsledky z funkce `regress`.

Jelikož model nevypadá zrovna dobře, zkuste ho upravit na model kvadratického trendu, tj. polynomu stupně dva (tj. je třeba změnit matici plánu a předpovědní matici plánu, dále všechny výrazy, kde vystupuje počet odhadovaných parametrů $k + 1$).

Příklad Na základě dat uvedených v souboru `hlavicka.m` se pokuste analyzovat závislost mezi velikostí hlavičky plodu a stářím dítěte jako nějaký polynomiální vztah, tj. vyjádřete stáří plodu jako funkci velikosti hlavičky plodu, pro polynomy různého stupně a na základě rozboru získaných údajů (významnost modelu, koeficient determinace, významnost parametrů, korelovanost reziduí ...) nalezněte nejvhodnější tvar této závislosti.

Z velikosti hlavičky plodu, kterou uložíte do proměnné `h`, určete délku těhotenství ve tvaru počet týdnu, počet dnů a také interval spolehlivosti pro tuto předpověď v témže tvaru. Určete pro velikosti hlavičky 7,5 cm a 8,5 cm. Jedná se o *interpolační předpověď*, tj. předpověď v bodě, ve kterém nebylo zachyceno pozorování. Vyzkoušejte též předpověď do budoucna, tj. *extrapolační předpověď*, pro ověření, zda je tento model vhodný pro předpovědi či naopak. Vyzkoušejte pro velikost hlavičky 24 cm.

5. Exponenciální trend

Jedná se o dvouparametrický trend tvaru:

$$Tr_t = \alpha \cdot \beta^t$$

kde $t = 1, \dots, n; \quad \alpha, \beta \in \mathcal{R}; \quad \beta > 0$

Podíl sousedních členů a také podíl sousedních diferencí je konstantní, rovný β. Je-li $\alpha > 0$, pak pro $\beta > 1$ roste a pro $0 < \beta < 1$ klesá.

Odhad parametrů provedeme logaritmizací (tj. převedeme na LRM):

$$\log Tr_t = \log \alpha + t \cdot \log \beta$$

tedy po přeznačení

$$Tr_t^* = b_0^* + t \cdot b_1^*$$

Jedná se tedy o model:

$$\boldsymbol{y}^* = \begin{pmatrix} \log y_1 \\ \vdots \\ \vdots \\ \log y_n \end{pmatrix} \quad \boldsymbol{X} = \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ \vdots & \vdots \\ 1 & n \end{pmatrix} \quad \boldsymbol{b}^* = \begin{pmatrix} \log \alpha \\ \log \beta \end{pmatrix}$$

Pokud byla náhodná složka v původním modelu přítomna aditivně a ne multiplikativně, vzniká problém (neadekvátní výsledky verifikace modelu), řešením je tzv. vážená regrese.

Dále se vážená regrese ukazuje vhodnější také proto, že logaritmování snižuje relativní váhu „velkých“ pozorování (tj. často nových) a zvyšuje relativní váhu „malých“ pozorování (tj. často starých).

Vážená regrese

Volba vah $w_t = y_t^2 \cdot v_t$, většinou se volí $v_t = 1$, tedy $w_t = y_t^2$.

$$W_{SSE} = \sum_{t=1}^{n} [w_t(\log y_t - \log \alpha - t \log \beta)]^2 \to \min$$

Jedná se tedy o model:

$$\begin{pmatrix} w_1 \log y_1 \\ \vdots \\ \vdots \\ w_n \log y_n \end{pmatrix} = \begin{pmatrix} w_1 & w_1 \cdot 1 \\ w_2 & w_2 \cdot 2 \\ \vdots & \vdots \\ w_n & w_n \cdot n \end{pmatrix} \cdot \begin{pmatrix} \log \alpha \\ \log \beta \end{pmatrix}$$

6. Modifikovaný exponenciální trend

Jedná se o zobecnění exponenciálního trendu ve tvaru:

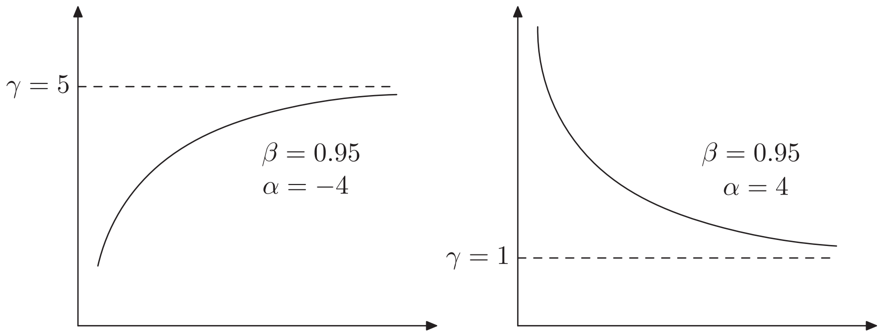
$$Tr_t = \gamma + \alpha\beta^t \qquad t = 1, \ldots, n; \quad \beta > 0$$

Jedná se o trend s konstatním podílem sousedních diferencí, tj.

$$\frac{\Delta Tr_t}{\Delta Tr_{t-1}} = \frac{Tr_t - Tr_{t-1}}{Tr_{t-1} - Tr_{t-2}} = \beta$$

Tento trend je asymptoticky omezen. Pro $0 < \beta < 1$ a $\alpha < 0$ roste, pro $\alpha > 0$ klesá, asymptotická úroveň γ se nazývá hladina saturace.

Tento model nelze linarizovat, používá se jiná metoda. Nevýhodou následující metody je nemožnost statistické verifikace parametrů a konstrukce funkcionálu f_t, který klademe rovný jedné. Soubor n pozorování rozdělíme na 3 stejně dlouhé části délky m a pozorování v jednotlivých částech sečteme.



Obrázek 6.3: Modifikovaný exponenciální trend

$$Tr_t = \gamma + \alpha\beta^t$$

$$\Sigma_1 = \sum_{t=1}^{m} y_t \quad \sim \quad m\gamma + \alpha\frac{\beta(\beta^m - 1)}{(\beta - 1)} = m\gamma + \alpha S$$

$$\Sigma_2 = \sum_{t=m+1}^{2m} y_t \quad \sim \quad m\gamma + \alpha\beta^m\frac{\beta(\beta^m - 1)}{(\beta - 1)} = m\gamma + \alpha\beta^m S$$

$$\Sigma_3 = \sum_{t=2m1}^{3m} y_t \quad \sim \quad m\gamma + \alpha\beta^{2m}\frac{\beta(\beta^m - 1)}{(\beta - 1)} = m\gamma + \alpha\beta^{2m} S$$

kde

$$S = \frac{\beta(\beta^m - 1)}{(\beta - 1)}$$

Odhad parametrů se provede následovně:

$$\Sigma_3 - \Sigma_2 = \alpha\beta^m S(\beta^m - 1)$$

$$\Sigma_2 - \Sigma_1 = \alpha S(\beta^m - 1) \quad \text{tj.}$$

$$\beta^m = \frac{\Sigma_3 - \Sigma_2}{\Sigma_2 - \Sigma_1} \quad \Rightarrow \quad \hat{\beta} = \left(\frac{\Sigma_3 - \Sigma_2}{\Sigma_2 - \Sigma_1}\right)^{\frac{1}{m}}$$

$$\hat{S} = \frac{\hat{\beta}(\hat{\beta}^m - 1)}{(\hat{\beta} - 1)}$$

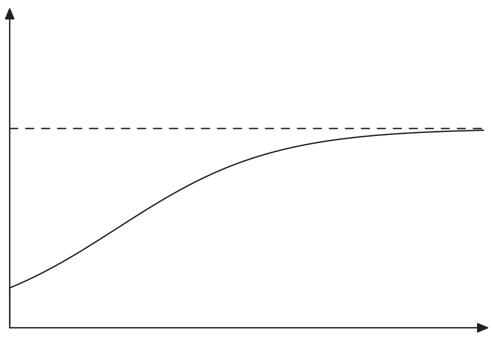
$$\Sigma_2 - \Sigma_1 = \alpha S(\beta^m - 1) \quad \Rightarrow \quad \hat{\alpha} = \frac{\Sigma_2 - \Sigma_1}{\hat{S}(\hat{\beta^m} - 1)}$$

$$\Sigma_1 = m\gamma + \alpha S \quad \Rightarrow \quad \hat{\gamma} = \frac{\Sigma_1 - \hat{\alpha}\hat{S}}{m}$$

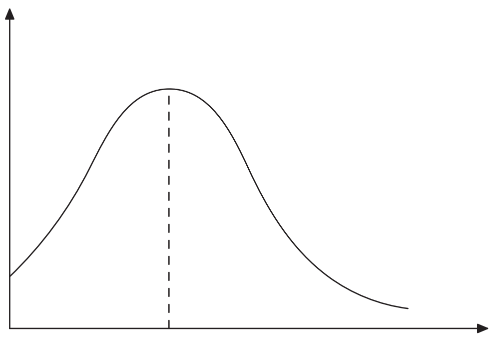
7. Logistický trend

- rostoucí funkce s inflexním bodem v čase $t = -\frac{ln(\alpha)}{ln(\beta)}$
- symetrická křivka
- často se užívá k popisu vybavenosti nějakým výrobkem - roste nejprve pozvolna, potom rychleji a jak se začíná přibližovat k hladině saturace γ, tak zase zpomaluje

$$Tr_t = \frac{\gamma}{1 + \alpha\beta^t} \qquad \alpha, \beta, \gamma \in \mathcal{R}, \quad \beta > 0, \quad t = 1, \ldots, n; \tag{7.8}$$



Obrázek 7.4: Logistický trend ($\alpha = 4$, $\beta = 0.95$, $\gamma = 5$)



Obrázek 7.5: Derivace logistické křivky

Odhad parametrů:

a) namodelujeme jako inverzi modifikovaného exponenciálního trendu

$$Tr_t = \frac{\gamma}{1 + \alpha\beta^t}$$

$$Tr_t^* = \frac{1}{Tr_t} = \frac{1}{\gamma} + \frac{\alpha\beta^t}{\gamma} = \gamma^* + \alpha^* \cdot \beta^t$$

kde $\gamma^* = \frac{1}{\gamma},\ \alpha^* = \frac{\alpha}{\gamma}$

b) nebo užijeme diferenční odhad parametrů

$$Tr_t = \frac{\gamma}{1 + \alpha\beta^t} \qquad \alpha, \beta, \gamma \in \mathcal{R}, \quad \beta > 0, \quad t = 1, \dots, n; \tag{7.9}$$

$$\frac{\partial Tr_t}{\partial t} = -\frac{ln\beta}{\gamma} y_t(\gamma - y_t) \tag{7.10}$$

$$\text{aproximace} \quad \frac{\Delta Tr_t}{\Delta t} = \frac{y_{t+1} - y_t}{(t+1) - t} = \Delta y_t \quad / : y_t \tag{7.11}$$

$$tr\_y_t \equiv \frac{\Delta y_t}{y_t} = -ln(\beta) + \frac{ln(\beta)}{\gamma} y_t \equiv -b_0^* + b_1^* y_t \tag{7.12}$$

Odhady parametrů b_0^* a b_1^* získáme MNČ.

$$tr\_y = \begin{pmatrix} tr\_y_1 \\ . \\ . \\ . \\ tr\_y_{n-1} \end{pmatrix} = \begin{pmatrix} \frac{\Delta y_1}{y_1} \\ . \\ . \\ . \\ \frac{\Delta y_{n-1}}{y_{n-1}} \end{pmatrix} \quad X = \begin{pmatrix} -1 & y_1 \\ . & . \\ . & . \\ . & . \\ -1 & y_{n-1} \end{pmatrix} \quad b = \begin{pmatrix} b_0^* \\ b_1^* \end{pmatrix}$$

Pozn.: $\Delta y_t = y_{t+1} - y_t$, proto jsou v maticích údaje jen po index $n-1$.

Původní parametry modelu:

$$\hat{\beta} = e^{\hat{b_0^*}} \qquad \hat{\gamma} = \frac{\hat{b_0^*}}{\hat{b_1^*}} \tag{7.13}$$

Pro odhad parametru α použijeme následující (Rhodesův) vztah:

$$\hat{\alpha} = \exp\left\{-\frac{(n+1)\ln(\hat{\beta})}{2} + \sum_{t=1}^{n} \frac{ln((\hat{\gamma}/y_t)-1)}{n}\right\} \tag{7.14}$$

8. Klouzavé průměry

Pokud se trend řady v čase mění a není možno jej proložit jednou matematickou křivkou, rozdělíme řadu na úseky a vyrovnáme každý zvlášť. Pokud se ovšem trend mění spojitě, je vhodné použít nějakou *adaptivní metodu* , tj. metodu, která se plynule přizpůsobuje změnám trendu. Je to např. metoda klouzavých průměrů či exponenciálního vyrovnání.

Máme n údajů a chceme jimi proložit klouzavé průměry délky $2m+1$: nejdříve zprůměrujeme prvních $2m+1$ hodnot a tuto hodnotu přiřadíme nějakému indexu (většinou tomu střednímu), potom se posuneme o pozorování doprava atd.

- Jednoduchý klouzavý průměr (nevážený)

  $$\hat{y}_t = \frac{1}{2m+1}(y_{t-m} + y_{t-m+1} + \ldots + y_{t+m})$$

  Problém: vyrovnání krajů a předpověď, lepší je

- Polynomiální klouzavý průměr (vážený)

  Polynomem stupně r postupně prokládáme $2m+1$ členů. Parametry odhadneme tak, aby byl součet čtverců chyb minimální. Posuneme se o pozorování doprava a postup opakujeme.

  Tedy hledáme parametry $b_0, b_1, \ldots, b_r$, které minimalizují:

  $$\sum_{\tau=-m}^{m} (y_{t+\tau} - (\hat{b_0} + \hat{b_1}\tau + \hat{b_2}\tau^2 + \ldots + \hat{b_r}\tau^r))^2 \to \min \tag{8.15}$$

  Tedy $\hat{y}_{t+\tau} = \hat{b_0} + \hat{b_1}\tau + \hat{b_2}\tau^2 + \ldots + \hat{b_r}\tau^r$. Vychází se z centrování okolo hodnoty y_t. Protože $\hat{y}_t = \hat{b_0} \ \ (\tau = 0)$, nemusíme nutně řešit všechny normální rovnice, které vzniknou optimalizací výrazu (8.15). Váhy jsou stejné pro klouzavé průměry řádu r a $r+1$, kde r je sudé. Váhy jsou symetrické.

Např. pro

$r = 2$ a $r = 3$, $2m + 1 = 5$: váhy $\frac{1}{35}(-3, 12, 17, \ldots)$

$r = 2$ a $r = 3$, $2m + 1 = 9$: váhy $\frac{1}{231}(-21, 14, 39, 54, 59, \ldots)$

$r = 4$ a $r = 5$, $2m + 1 = 9$: váhy $\frac{1}{429}(15, -65, 30, 135, 197, \ldots)$

Pozor na kraje a předpověď!!! To se dělá jinak.

Metoda umožňuje vyrovnat i všechny krajní hodnoty i konstrukci předpovědi. K tomu je však třeba odhadnout všechny neznámé parametry. Váhy již nejsou stejné pro dva řády, ani nejsou symetrické !!!!! Předpověď nemá smysl dělat na příliš mnoho kroků do budoucnosti.

Pro $r = 2$ a $r = 3$, $2m + 1 = 5$:

$\hat{y_1} = \frac{1}{70}(69, 4, -6, 4, -1)$

$\hat{y_2} = \frac{1}{35}(2, 27, 12, -8, 2)$

$\hat{y_{n-1}} = \frac{1}{35}(2, -8, 12, 27, 2)$

$\hat{y_n} = \frac{1}{70}(-1, 4, -6, 4, 69)$

$\hat{y_{n+1}} = \frac{1}{5}(-4, 11, -4, -14, 16)$

Alternativou k dosazení vah, které najdeme v tabulkách, je přímé užití lineární regrese. Sestavíme matici plánu a odhadneme všechny parametry pro každou skupinu $2m + 1$ pozorování. Není ovšem možné verifikovat parametry (příliš málo stupňů volnosti způsobuje, že parametry vycházejí statisticky nevýznamné). Ani model jako celek nemá smysl verifikovat.

větší r = přesnější proložení

delší 2m+1 = větší vyhlazení

Existují objektivní kriteria pro volbu délky a řádu klouzavého průměru (problémy se zahrnutím cyklické složky). Metoda způsobuje autokorelaci reziduí.

9. Exponenciální vyrovnání

Tato metoda se snaží křivkou vyrovnat vždy všechna minulá pozorování. Přitom se předpokládá, že význam pozorování do minulosti exponenciálně klesá (starším pozorováním přiřadíme exponenciálně klesající váhy). Minimalizujeme tedy kriterium:

$$SSE = \sum_{\tau=0}^{t-1} \alpha^{\tau} (y_{t-\tau} - \hat{y}_{t-\tau})^2 \rightarrow \min \tag{9.16}$$

kde $\alpha \in \langle 0;\ 1 \rangle$ je tzv. koeficient zapomínání.

Tato metoda odhaduje parametry pro každý časový okamžik zvlášť, parametry se tedy adaptují v čase.

Podle toho, zda k vyrovnání použijeme polynom řádu 0, 1 či 2, hovoříme o jednoduchém, dvojitém či trojitém exponenciálním vyrovnání.

9.1 Jednoduché exponenciální vyrovnání

Předpokládáme, že se trend v krátkém období vyvíjí zhruba konstantně. Vyrovnaná hodnota v čase t na základě informací dostupných v čase t, je pak rovna odhadu parametru polynomu řádu nula, stejně tak předpověď.

$$\hat{y}_t = \hat{y}_t(t) = \hat{b}_0(t); \quad \hat{y}_{t+k}(t) = \hat{b}_0(t) \tag{9.17}$$

Parametr $\hat{b}_0$ minimalizuje kriterium

$$SSE = \sum_{\tau=0}^{t-1} \alpha^{\tau} (y_{t-\tau} - \hat{b}_0(t))^2 \rightarrow \min \tag{9.18}$$

Parametr $\hat{b}_0$ vypočteme následovně:

$$\hat{b}_0(t) = (1 - \alpha) \cdot y_t + \alpha \cdot \hat{b}_0(t-1) \tag{9.19}$$

Počáteční hodnotu parametru $\hat{b}_0(0)$ odhadneme jako průměr několika prvních pozorování či jako průměr celé řady. Jako vhodný koeficient zapomínání α zvolíme ten, který dává nejměnší součet čtverců chyb předpovědi. Pokud je to méně než 0.7, není tato metoda zřejmě na data vhodná.

Předpovědní interval spolehlivosti má tvar:

$$(\hat{y}_{t+\tau} - u_{1-\alpha/2} \cdot d_\tau \cdot \Delta(n);\ \hat{y}_{t+\tau} + u_{1-\alpha/2} \cdot d_\tau \cdot \Delta(n)) \tag{9.20}$$

kde

$$d_\tau = 1.25; \quad \Delta(n) = \sum_{t=1}^{n} \frac{|y_t - \hat{y}_t(t-1)|}{n} \tag{9.21}$$

9.2 Dvojité exponenciální vyrovnání

Použijeme, pokud se daná veličina vyvíjí v krátkém období zhruba lineárně.
Parametry $\hat{b}_0$ a $\hat{b}_1$ minimalizují kriterium

$$SSE = \sum_{\tau=0}^{t-1} \alpha^\tau (y_{t-\tau} - \hat{b}_0(t) - \hat{b}_1(-\tau))^2 \to \min \tag{9.22}$$

1. fáze - hledání optimální hodnoty α

- nejprve provedeme regresní odhady počátečních hodnot parametrů, tj. $b_0(0)$ a $b_1(0)$, z prvních 6 nebo $n/2$ hodnot
- dále pro každé α spočteme vztahy (9.23) – (9.28)
- zvolíme to α, pro které je suma chyb (9.28) minimální

Počáteční hodnoty jednoduché a dvojité vyrovnávací statistiky:

$$S_0 = b_0(0) - \frac{\alpha}{1-\alpha} b_1(0) \tag{9.23}$$

$$S_0^{[2]} = b_0(0) - \frac{2\alpha}{1-\alpha} b_1(0) \tag{9.24}$$

Jednoduchá vyrovnávací statistika v čase t

$$S_t = (1-\alpha)y_t + \alpha S_{t-1} \qquad t = 1, 2, \dots \tag{9.25}$$

Dvojitá vyrovnávací statistika v čase t

$$S_t^{[2]} = (1-\alpha)S_t + \alpha S_{t-1}^{[2]} \qquad t = 1, 2, \dots \tag{9.26}$$

Předpověď na čas $t+\tau$ na základě informací dostupných v čase t

$$(*) \quad \hat{y}_{t+\tau}(t) = \hat{b}_0(t) + \hat{b}_1(t)\tau = \left(2 + \frac{1-\alpha}{\alpha}\tau\right) S_t - \left(1 + \frac{1-\alpha}{\alpha}\tau\right) S_t^{[2]}$$

Tedy předpověď na čas $t+1$ na základě informací dostupných v čase t

$$\hat{y}_{t+1}(t) = \left(2 + \frac{1-\alpha}{\alpha}\right) S_t - \left(1 + \frac{1-\alpha}{\alpha}\right) S_t^{[2]} \tag{9.27}$$

Součet čtverců chyb jednokrokových predikcí

$$\text{suma chyb} = \sum_{t=0}^{n-1} (\hat{y}_{t+1}(t) - y(t+1))^2 \to \min \tag{9.28}$$

2. fáze (vlastní exponenciální vyrovnání)

- provedeme počáteční regresní odhady parametrů, $b_0(0)$ a $b_1(0)$, tentokráte z celé řady
- spočteme hodnoty S_0, $S_0^{[2]}$, S_t, $S_t^{[2]}$ a $\hat{y}_{t+1}(t)$ dle vztahů (9.23), (9.24), (9.25), (9.26) a (9.27)
- dále spočteme vyrovnané hodnoty (do vztahu $(*)$ dosadíme $\tau = 0$), budoucí předpověď a interval spolehlivosti dle vztahů (9.29), (9.30), (9.31) a (9.32)

Vyrovnané hodnoty (tj. predikce pro $\tau = 0$)

$$yv \equiv \hat{y}_t(t) = 2S_t - S_t^{[2]} \tag{9.29}$$

Konfidenční interval předpovědi

$$\mathrm{IS} = (\hat{y}_{t+\tau} - u_{1-\alpha/2} \cdot d_\tau \cdot \Delta(n);\ \hat{y}_{t+\tau} + u_{1-\alpha/2} \cdot d_\tau \cdot \Delta(n)) \tag{9.30}$$

kde

$$\Delta(n) = \sum_{t=1}^{n} \frac{|y_t - \hat{y}_t(t-1)|}{n} \tag{9.31}$$

$$d_\tau = 1.25\sqrt{\frac{1 + d_1(d_2 + d_3 d_4 \tau + d_5 \tau^2)}{1 + d_1(d_2 + d_3 d_4 + d_5)}} \tag{9.32}$$

kde

$$d_1 = \frac{1-\alpha}{(1+\alpha)^3};\quad d_2 = 1+4\alpha+5\alpha^2;\quad d_3 = 2(1-\alpha);\quad d_4 = 1+3\alpha;\quad d_5 = 2(1-\alpha)^2;$$

10. Sezónní složka

10.1 Jednoduchý přístup

Pomocí některé z dříve uvedených metod odhadneme trend řady $\hat{Tr}_t$ a u aditivního modelu jej odečteme od dat y_t (resp. vydělíme jím u multiplikativního modelu data y_t).

Zbyde nám tedy sezónní složka a náhodná složka. Máme-li m sezón za období, vezmeme každou m-tou složku vektoru $y_t - Tr_t$ resp. y_t/Tr_t a zprůměrujeme je, tento postup opakujeme m-krát (např. pro $m = 4$ opakujeme pro 1., 2., 3., a 4. čtvrtletí). Vyrovnané hodnoty jsou pak součtem (resp. součinem pro multiplikativní model) odhadnutého trendu a odhadnuté sezónní složky.

První vztah vždy platí pro aditivní model, druhý pro multiplikativní model.

$$S_t + E_t = y_t - \hat{Tr}_t \quad \text{resp.} \quad S_t \cdot E_t = y_t/Tr_t$$

Požadavek $\sum \hat{S}_i = 0$ resp. $\sum \hat{S}_i = m$, kde m je počet sezón.

10.2 Regresní přístup

$$S_t = b_1 x_{t1} + \ldots + b_m x_{tm}$$

x_{it} - umělé proměnné, $x_{it} = \begin{cases} 1 & \text{když } t = \text{sezóně } i \\ 0 & \text{jinak} \end{cases}$

Sezónní část matice plánu pro čtvrtletní a měsíční údaje (tzv. regresní hřeben):

$$XS_4 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \end{pmatrix} \quad XS_{12} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix}$$

Pokud bychom chtěli odhadovat zároveň trend i sezónu, regresní rovnice by vypadala takto:

$$y_t = Tr_t + S_t + u_t = \underbrace{b_0 + b_1 x_{t1} + \ldots + b_k x_{tk}}_{Tr_t} + \underbrace{b_{k+1} x_{t,k+1} + \ldots + b_{k+m} x_{t,k+m}}_{S_t} + u_t$$

V tomto případě by však byl první sloupec matice plánu (samé jedničky) lineární kombinací sloupců sezónní části matice plánu, tudíž by byl porušen předpoklad pro použití MNČ říkající, že hodnost matice plánu musí být rovna počtu jejích sloupců. V praxi se tedy jeden sloupec (jedničky nebo sloupec u první sezónní proměnné) vynechává, úrovňová konstanta se pak dopočte zvlášť. Tedy např. pro kvadratický trend a čtvrtletní údaje vypadá matice plánu takto:

$$X = \begin{pmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 2 & 2^2 & 0 & 1 & 0 & 0 \\ 1 & 3 & 3^2 & 0 & 0 & 1 & 0 \\ 1 & 4 & 4^2 & 0 & 0 & 0 & 1 \\ 1 & 5 & 5^2 & 1 & 0 & 0 & 0 \\ 1 & 6 & 6^2 & 0 & 1 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix}$$

Při zadávání do regrese zadáme místo X např. `X(:,2:end)` nebo `X(:,[1:3,5:7])`.

11. Ekonometrický model

Na rozdíl od dekompozičních modelů se předpokládá, že sledovaná veličina není funkcí času, ale určitých vysvětlujících proměnných na základě ekonomické hypotézy. Např. keynesiánská či Friedmanova spotřební funkce apod. Ekonometrické modely je třeba testovat na multikolinearitu.

$$C_t = \overline{C} + c \cdot Y_t + u_t$$

$$C_t = b_1 Y_t + b_2 Y_{t-1} + \ldots + b_k Y_{t-k+1} + u_t$$

$$C_t = \overline{C} + b_1 \cdot Y_{t-1} + b_2 \cdot C_{t-1} + b_3 \cdot r_t + u_t$$

11.1 Multikolinearita

Je-li porušen předpoklad vzájemné nezávislosti vysvětlujících proměnných, tj. je-li matice plánu $X_{n \times (k+1)} \Rightarrow h(X) < k+1 \Rightarrow |X'X| = 0$ a nelze užít pro odhady parametrů MNČ. Hovoříme o *multikolinearitě* , tj. o úplné lineární závislosti vysvětlujících proměnných.

Pokud se determinant $|X'X|$ k nule pouze blíží, nastávají problémy s numerickými výpočty, jedná se o částečnou lineární závislost vysvětlujících proměnných, tj. částečná multikolinearita.

Testování multikolinearity

normování proměnných:

$$x^0_{ti} = \frac{x_{ti} - \overline{x}_{ti}}{\text{std}(x_{ti})}$$

korelační matice:

$$R = \frac{1}{n}((X^0)' \cdot X^0)$$

kde matici X^0 tvoří sloupce x_{ti}^0

$|R| \in \langle 0, 1 \rangle$:

$|R| = 1$ - žádná lineární závislost

$|R| = 0$ - úplná lineární závislost mezi dvěma a více vysvětlujícími proměnnými

$|R| \to 0$ - multikolinearita je významnější

Metoda Farrara a Glaubera

$$\chi_R^2 = -(n - 1 - \frac{2k+5}{6}) \cdot \log |R| \sim \chi^2(\nu = k(k-1)/2)$$

Pokud $\chi_R^2 > \chi_{1-\alpha}^2(\nu)$, pak s pravděpodobností $1-\alpha$ zamítáme hypotézu o nepřítomnosti multikolinearity.

Co s tím

Zjistíme, která proměnná multikolinearitu způsobuje. Tuto veličinu pak buď z modelu vypustíme či nějak transformujeme (centrování, normování). Spočteme:

$r^{ii} = \frac{|R_{-1}|}{|R|}$ kde R_{-1} vznikne z R vypuštěním i-tého sloupce

$F_i = (r^{ii} - 1) \cdot \frac{n-k}{k-1}$

Pokud $F_i > F_{1-\alpha}(k-1, n-k)$, pak s pravděpodobností $1-\alpha$ zamítáme hypotézu o nekorelovanosti veličiny x_i s ostatními veličinami.

11.2 Koeficienty beta

Říkají, jak se která vysvětlující proměnná podílí na vysvětlení vysvětlované veličiny y_t. Je-li např. $\beta_i = 0.3$, znamená to, že i-tá proměnná vysvětluje 30% chování dat y_t.

$$\beta_i = \frac{|cor(y_t, x_{ti})|}{\sum_{j=0}^{k} |cor(y_t, x_{tj})|}$$

kde

$$cor(y_t, x_{ti}) = \hat{b}_i \cdot \frac{\sqrt{\frac{1}{n}\sum_{t=1}^{n}(x_{ti} - \overline{x}_i)^2}}{\sqrt{\frac{1}{n}\sum_{t=1}^{n}(y_t - \overline{y})^2}}$$

11.3 Příklad – model poptávky po penězích

Vyjdeme z jednoduchého Baumol-Tobinova modelu poptávky po penězích, který má tvar

$$\frac{M^d}{P} = \sqrt{\frac{pT}{2i}} \tag{11.33}$$

kde M^d je poptávané množství peněz, P je cenová hladina, T je reálný objem transakcí, i je nominální úroková míra a p je konstanta (náklady na prodej obligací).

Logaritmováním převedeme rovnici (11.33) do tvaru

$$\log M^d - \log P = \log \sqrt{T} + \log \sqrt{p} - \log \sqrt{2} - \log \sqrt{i} \tag{11.34}$$

Dále nahradíme $\log \sqrt{p} - \log \sqrt{2}$ jedním parametrem b_0, parametrizujeme vysvětlující proměnné a nahradíme nepozorovaný objem transakcí T pozorovaným údajem o reálném hrubém domácím produktu Y. Pokud umístíme všechny vysvětlující a vysvětlované proměnné do stejného času t, získáme tím pro jednotlivá období $t = 1, \ldots, n$ soustavu stochastických rovnic

$$(\log M_t - \log P_t) = b_0 + b_1 \log \sqrt{Y_t} + b_2 \log \sqrt{i_t} + \epsilon_t \tag{11.35}$$

neboli po přeznačení

$$(m_t - p_t) = b_0 + b_1 y_t + b_2 r_t + \epsilon_t \tag{11.36}$$

kde b_i jsou neznámé hodoty parametrů (platí, že $b_1 > 0$ a $b_2 < 0$) a ϵ_t je náhodná složka.

Protože pozorované hodnoty HDP (hodnoty Y_t), cenového indexu (P_t) a úrokové míry (hodnoty i_t) jsou sezónně očištěné, zatímco údaje o vývoji peněz v oběhu ne, je rovnice (11.35) zatížena výraznou sezónností. Proto je třeba vytvořit umělé proměnné $u_{1t}, \ldots, u_{4t}$, pomocí kterých bude ošetřen vliv sezónní složky (pozorování jsou čtvrtletní). Do modelu dáme jen tři z těchto umělých proměnných nebo vypustíme úrovňovou konstantu, abychom v modelu nezpůsobili multikolinearitu.

Komentář k jednotlivým krokům verifikace modelu je součástí souborů Program1.m, Program2.m a Program3.m. Data jsou uložena v souboru `dataset.m`. `Program1.m` je odhadem rovnice (11.36), `Program2.m` je rozšířen o odhad sezónní složky, `Program3.m` obsahuje pokus o předpoveď do budoucna.